Fama-French Models Application to the analysis of FTSE4GOOD, MSCI ESG and STOXX ESG Indices:

Do SRI Indices have biases?

by

Pedro Gil Gonçalves Lima

Dissertation on Master in Economics and Business Administration

Supervisor:

Abel Luís da Costa Fernandes

2017

Bibliographical Note

The candidate was born on 21st February 1993 on Aveiro. On 2014, he simultaneously obtained a Bachelor in Economics in the University of Aveiro and concluded studies in Piano at the Conservatory of Music of Aveiro Calouste Gulbenkian.

Afterwards, instead of enrolling immediately in a Master Degree, he opted for a gap year to reflect better upon his future as well to pursue other passions, such as continuing to study Piano and learning German and Chinese.

After enrolling in the Master in Economics and Business Administration, he continued to pursue in parallel those same passions. He gives tutoring lessons on both areas of economics and music because he loves to teach and even more to learn.

Acknowledgements

First of all, I would like to express my deepest thanks to my supervisor, Professor Abel Fernandes, for his encouragement since the beginning, his effort to answer my questions and doubts as they arose and helpful guidance and commentaries without which it would be difficult to successfully complete this Dissertation.

To my family, in particular my parents and my brother for their patience and support throughout this journey.

I also would like to thank my teachers who, during my time as a student, contributed, each in their own way, to my growth as an individual and by the impact they had, sometimes even with just a few words.

And last but not least to my friends and colleagues who accompanied me throughout this process.

In particular, I would like to thank Daniela for being who she is and for all the support she has given me, Teresa for her friendship and kindness and André for his companionship.

Abstract

We apply Fama-French models to the analysis of Socially Responsible Investment Indices from FTSE4GOOD, MSCI ESG and STOXX ESG Series. We also apply novel approaches to the analysis of such indices, such as adjusting for industry effects, applying a GMM-System framework and comparing between crisis and non-crisis periods. Most literature on Socially Responsible Investment has been focused on analysing Funds rather than Indices. However, there's an increasing amount of evidence that questions their suitability to act as a proxy for this type of Investment, namely failures in keeping ethical standards. Moreover, the lack of application of Fama-French models to SRI Indices leaves room for a more detailed analysis of what biases these Indices may have and their consequences for risk-adjusted performance. Overall, SRI Indices provide a very similar risk-adjusted performance as their conventional benchmarks and tend to be biased towards loser stocks. We also detected the presence of industry effects. During crises, there's a non-statistically significant increase in performance and SRI Indices tend to be less risky and more exposed to Value Companies. However, overall, there are multiple differences between indices both at the country and at the Series level which highlights the importance of analysing multiple sources of SRI to obtain more representative results. Global Indices tend to have higher risk than regional Indices and FTSE4GOOD Indices tend to be more biased towards large companies with weak profitability profiles while indices from other series tend to show a small cap bias and no pattern in terms of profitability exposure. Moreover, results are sensitive to the application of more complex procedures, namely the sign of alpha as well as value and investment exposures.

Keywords: Socially Responsible Investment, FTSE4GOOD, MSCI ESG, STOXX ESG, biases, Fama-French models, crisis, industry factors, GMM System.

Resumo

Nesta tese, fazemos uma aplicação de modelos Fama-French à análise de Indices de Investmento Socialmente Responsável das séries FTSE4GOOD, MSCI ESG e STOXX ESG. Aplicam-se igualmente outras abordagens inovadoras a Índices SRI tais como análises ajustadas a enviesamentos industriais, a aplicação de Sistemas GMM e a análise de crises. A maior parte da literatura tem-se concentrado na análise de Fundos em detrimento de Índices. No entanto, muitos autores põem em causa a sua adequabilidade para representar este tipo de investmento, nomeadamente devido a falhas na manutenção de padrões éticos. Acresce que a falta de aplicação de modelos Fama-French à análise deste índices deixa espaço para uma análise mais detalhada de que enviesamentos estes podem ter assim como as suas consequências em termos de performance ajustada ao risco. Globalmente, os índices apresentam uma performance muito semelhante em relação aos seus benchmarks oficiais e tendem a ter um enviesamento para ações perdedoras. Também detetámos a presença de enviesamentos relacionados com indústrias. Durante crises, existe um aumento não estatisticamente significativo de performance e uma tendência para menor risco e maior exposição a ações de valor. No entanto, existem muitas diferenças entre índices quer de diferentes países quer de diferentes séries o que realça a importância de fazer análises que tenham em conta múltiplas fontes deste tipo de investimento para chegar a resultados mais representativos. Algumas das diferenças principais são o facto de Índices Globais apresentarem um maior risco em relação a Índices Regionais e Índices da FTSE4GOOD tenderem a ser mais enviesados para empresas grandes e com fracos lucros em relação às outras Séries, que estão mais investidas em empresas pequenas e sem padrões em termos de lucro. Alguns resultados tendem a variar consoante o modelo aplicado, nomeadamente o sinal do Alpha bem como as exposições ao fator valor e investimento.

Palavras-chave: Investimento Socialmente Responsável, FTSE4GOOD, MSCI ESG, STOXX ESG Leaders, enviesamentos, Modelos Fama-French, crise, Fatores índustriais, Sistema GMM.

List of Abbreviations

AR – Absolute Return;

CFP – Corporate Financial Performance;

CSP – Corporate Social Performance;

CSR – Corporate Social Responsibility;

ESG – Environment, Social, Governance;

ER – Excess Return;

EU – European Union;

FTSE – Financial Times Stocks Exchange;

F-F – Fama-French;

GMM – Generalized Method of Moments;

OLS – Ordinary Least Squares;

MSCI – Morgan Stanley Capital International;

NA – North America;

RAP – Risk-Adjusted Performance;

SRF – Socially Responsible Funds;

SRI – Social Responsible Investment;

SUR – Seemingly Unrelated Regressions;

TR – Total Return;

UK – United Kingdom;

US – United States.

Index

Index of Tables

Index of Figures

1. Introduction

Socially Responsible Investment (hereafter SRI) are investments that take into account ethical, moral and other criteria besides financial performance (Barnett and Salomon, 2006; Becchetti et al., 2015; Bello, 2005). This type of investment began mainly in sixties and seventies and due to an increasing interest from private and institutional investors experienced an enormous growth over the years and have since expanded into Europe, United States and, more recently on Asia (Managi, 2012).

Most authors have reached different conclusions about the relative performance of SRI Vehicles in relation to conventional Investments, but most results point to non-statistically significant differences between the two types of investment. However, most have analysed Funds instead of Indices, which may not the most appropriate option. There's an increasing amount of evidence that questions their suitability to represent SRI, namely due to lack of ethical nature and transparency. Moreover, few authors have studied SRI Indices and applied Fama-French models, which allow to take into account possible biases that this type of investment may imply in relation to conventional investments. Thus, their application to the study of SRI Indices is necessary to understand the main differences between companies that are included in a SRI Index and their impact in terms of performance.

We apply Fama-French models to the analysis of SRI Indices from three Index Series, FTSE4GOOD, MSCI ESG and STOXX ESG. In this analysis, we also apply procedures which, to the best of our knowledge, were not yet applied to the analysis of SRI Indices, such as the addition of industry factors, the application of a Generalized Method of Moments-system framework and the comparison between crisis and non-crisis periods. Our analysis is also more consistent and representative of SRI by choosing Indices with similar geographic scope from three different SRI providers.

Thus, we provide robust answers to the following questions:

- What is the performance of SRI Indices in relation to conventional Investments?
- Are SRI Indices biased?
- What patterns exist in terms of performance and risk-exposures when comparing different Indices from different SRI Series?
- What patterns exist in terms of performance and risk exposures when analysing SRI Indices with similar geographic Scope?
- How does the performance and risk-exposures of SRI Indices vary between crisis and non-crisis periods?

The remainder of this Thesis is structured in the following manner:

- Chapter 2: Review of the Literature – We analyse more than 80 papers from SRI literature and expose our main findings regarding four topics: Structure, Performance, Biases, Time-varying performance and risk-exposures. We conclude with a summary of our findings and the main Contributions we provide to the literature.
- Chapter 3: Data and Methodology – We provide Descriptive statistics and a description of the main features of the screening process of the FTSE4GOOD, MSCI ESG and STOXX ESG Series. We also describe in detail the models, estimation procedures applied in their analysis, robustness tests and limitations of our methodology.
- Chapter 4: Discussion of Results – We made a detailed analysis of the results obtained for each model.
- Chapter 5: Conclusions and Directions for Future Research – We summarise the results we obtained, explain their main implications and suggest new avenues of research.

2. Review of the Literature

Socially Responsible Investment (hereafter SRI) are investments that take into account ethical, moral and other criteria besides financial performance (Barnett and Salomon, 2006; Becchetti et al., 2015; Bello, 2005).This type of investment is also called Ethical Investment (Bauer et al., 2007; Blanchett, 2010), Sustainable Investment or Conscious Investment (Blanchett, 2010) but Socially Responsible Investment is the most usual definition (Dam and Scholtens, 2016; Wallis and Klein, 2015).

We made a thorough analysis of more than 80 empirical papers[1] from this field whose individual findings are summarised in tables present in Appendix 1. We structure this section into the following subsections:

- Structure – We analyse the structure of the literature, which can be divided into three assets levels of analysis, Funds, Stocks and indices. However, most papers belong to the first category which can pose some problems that we describe in detail.
- Performance – We present an overall view of findings in the literature and describe the main approaches used by authors to assess performance. Moreover, we describe patterns we found related to certain topics such as Countries, Time periods, Dimensions and Diversification.
- Biases – We make an analysis of the various biases discovered by authors that have applied Fama-French based models to SRI vehicles. Moreover, other biases related to industries and local factors are also analysed.
- Time-varying Performance and risk exposures – We describe in detail the various approaches used to deal with the time-varying nature of performance measures and risk exposures, such as applying conditional approaches and analysing crisis and non-crisis periods.

[1] Besides empirical works, in some sections we also mention authors that made different studies related to SRI and which are relevant in some way.

- Summary – We summarise the most important findings from the Review of the literature.
- Contributions – Based on our findings we summarize our main contributions we provide to the literature.

2.1 Structure

SRI literature can be divided into three assets levels of analysis: funds, stocks and indices. The first category is the most common one (41 papers) and consists in comparing Socially Responsible Investment Funds (hereafter SRI Funds) with conventional benchmarks.[2] The second one (17 papers) implies obtaining stocks from a Database as well their respective ESG scores[3], constructing synthetic portfolios[4] and compare high[5] and low rated portfolios. The third category (16 papers) consists in comparing Socially Responsible Investment Indices (hereafter SRI Indices) with a conventional benchmark which typically is its official benchmark. Within this group, as can be seen in Table 17, the most analysed indices are FTSE4GOOD, DJSI and DSI 400 Indices but other lesser-known Indices are also studied. Finally, some authors opted for more than one level of analysis (7 studies) and the choice often falls on analysing funds and indices.[6]

The fact that most papers in the literature have analysed SRI at the fund level can pose a problem as many authors have questioned their ethical nature. Schwartz (2003) found that SRI funds were not sufficiently meeting their obligations namely in terms of complete information disclosure and non-deceptive advertising. This lack of transparency was also noticed by other authors that made empirical studies. Martí-Ballester (2003) was not able to determine the nature and quantity of screenings[7] applied by funds included in their sample. Scholtens et al. (2005) refrained from including a CSR[8] variable in their model despite recognizing its possible usefulness because information was limited to yearly reports and was only stored during three years. Cortez et al. (2012) discovered a large cap bias in funds that

---

[2] SRI Funds are identified by reliable sources such as Morningstar, Lipper, Bloomberg and Social Investment Forum which provide lists used by many authors to obtain their samples.
[3] "ESG scores" is a known term which appears often in SRI literature and refers to scores of companies on specific dimensions, namely Environment, Social and Governance (ESG) criteria. Examples of agencies that provide such ratings include KLD, Vigeo, SAM and others.
[4] Term used by authors such as Schröder (2007) since it refers to artificial portfolios that don't actually exist.
[5] High (low) rated portfolios are portfolios comprised with stocks with high (low) scores on some particular dimension.
[6] One exception was Vermeir (2005) that chose indices and stocks instead.
[7] Screenings are filters used to select what companies are included in a portfolio. They are based on certain criteria that will be further explained in the section "Performance".
[8] Corporate Social Responsibility. Other terms will be used interchangeably such as Corporate Social Performance (CSP) as many authors have also applied this term.

were labelled small cap and a growth bias in funds labelled value. This type of error of classification casts doubt on the reliability and transparency of labels of funds.

A recent study of Wimmer (2013) found that ESG scores of SRI Funds no longer persisted after three years which implies that after some time they tend to neglect their ethical commitments. Rhodes (2010) points out that SRI fund managers have difficulties in defining, applying and confirming adherence to screening criteria. Some authors found that returns of SRI Funds were more correlated with a conventional benchmark than with SRI Benchmarks (Bauer et al., 2005; Bauer et al., 2007; Climent and Soriano, 2011; Leite et al., 2014) which casts serious doubts on their ethical nature. Renneboog et al. (2008) found similar results using a slightly different approach. In particular, the authors found that the addition of an ethical benchmark to a Carhart model (1997) didn't contribute to increase the explanatory power of the model.

The analysis of funds may also be influenced by multiple factors not related to the application of social screenings such as management fees, transaction costs and other characteristics which are specific features of funds[9] (Sauer et al., 1997). Most authors try to avoid these pitfalls by applying a matching procedure in the construction of their sample. For each SRI Fund, they select conventional funds with similar characteristics. Typically, the matching criteria are inception date, total net assets and investment objective (see for e.g. Leite et al. (2015), Munöz et al. (2014); Nofsinger and Varma (2014)) or risk exposures[10] (Becchetti et al., 2015). However, this procedure is complex and can easily miss some important feature (Schröder, 2007) and these differences become more evident as different countries are analysed, which is the case of many papers (Rehman et al. 2016).

The distinction between SRI Funds and Conventional Funds is gradually becoming less clear. For instance, Borgers et al. (2015) showed that while SRI Funds have very small exposure to socially sensitive sectors, the same is true with Conventional Funds. A recent work of Duuren et al. (2016) concluded that there was growing tendency for Conventional Funds

[9] Thus, analysis at other levels (index and stock) don't suffer from this kind of problem.
[10] Risk exposures includes not only exposures to the market (CAPM) but also to other risk factors such as size, value (Fama-French, 1993) and momentum (Carhart, 1997).

Managers to adopt an investment process that takes into account typical concerns of responsible investment processes.

Since these problems are specific to SRI Funds, analysis at the stock or index level may be ore adequate. However, in the case of stocks, it involves constructing artificial portfolios that don't actually exist. Indices, on the other hand, represent well-known portfolios that serve as a guideline to both private and institutional investors (Schröder, 2007). Moreover, the only difference between a SRI Index and its official benchmark is the application of screening criteria (Ortas et al., 2013). Thus, composition only changes in response to social and ethical concerns and not to other factors (Sauer et al., 1997). The construction methodology of indices is also much more transparent and balanced as several stakeholders, such as researchers, NGO's and international agencies have been shown to have an increasingly important role in the definition of selection criteria of many indices (Fowler and Hope, 2007).

These advantages make the analysis of indices more interesting, especially considering that few authors have studied them (Fowler and Hope, 2007; Rehman et al., 2016; Wallis and Klein, 2014) and that they have showed a significant development in the last decade, especially since 2006 (Sun et al., 2011).

2.2 Performance

The literature on SRI has grown exponentially over the years, with the main focus on determining if it was possible to "do well while doing good"[11], i.e., being able to make investments that simultaneously attended to environmental, social, governance and similar ethical concerns and had competitive financial performance.

Theoretically, Hamilton (1993) developed three hypotheses for the relative performance of SRI in relation to conventional investments. The first hypothesis is that there is no difference between the risk-adjusted performance of SRI and conventional investments. This implies that the socially responsible component of an investment is not priced by the market. The second hypothesis is that the expected return of SRI is lower than Conventional Investments. SRI Investors contribute to increase the value of the company by expecting a lower return in relation to Conventional Investments. The third hypothesis is that the expected return of SRI surpasses those of Conventional Investments because Conventional Investors constantly underestimate the probability and impact of the release of negative public information about companies that are not socially responsible.

By looking at the empirical literature, we can see a great variety of results, which doesn't provide support for any particular hypothesis. This is the case even when looking at more recent literature. Positive results[12] are becoming rarer but still appear (Hooi et al., 2015; Lins et al., 2016) but findings still range from negative (Charfeddine et al., 2016; Silva and Cortez, 2016) to non-statistically significant results (Lesser et al., 2016; Rehman et al., 2016). Many authors obtain mixed findings (Auer et al., 2016; Becchetti et al., 2015; Erragragui and Revelli, 2016).

As can be seen in the column "Methodology" of the tables in Appendix 1, in order to assess performance, many approaches have been followed. However, essentially, they can be

[11] Expression used by Hamilton (1993).
[12] Since most authors have used benchmark models such as CAPM, Fama-French (1993) and Carhart (1997) to evaluate the financial performance of SRI relative to conventional investments, to avoid needless repetition we will often use interchangeably the terms "alpha" with "performance", "results", "findings" and similar expressions.

summed up in three categories. The first is the most popular one and consists of obtaining the Jensen's alpha (Jensen, 1968) from benchmark models such as CAPM, Fama-French (1993) and/or Carhart (1997). The CAPM involves regressing excess returns[13] of the SRI Vehicle on the excess return of the market. In the case of Fama-French (1993) model, the explanatory variables also include self-financing portfolios intended to reproduce additional risk factors related to size and value effects. The Carhart (1997) model is based on the previous model and adds a self-financing portfolio intended to capture the momentum anomaly detected by Jegadeesh and Titman (1993). More recently, Fama-French (2015) added profit and investment portfolios and some authors have used this models in SRI literature. The alpha represents the difference between the effective and expected return of the asset according to the used model[14]. Since the market is typically represented by a conventional benchmark, the alpha indicates if the SRI vehicle outperformed, underperformed or showed similar performance to conventional investments.[15]

The second approach involves estimating and comparing risk-adjusted performance measures of SRI and conventional investment vehicles. The Sharpe Ratio is the most popular measure and is defined as the ratio of the excess return of the asset and the Total risk, which is measured by the tracking error[16]. The Treynor Ratio is defined as the ratio of the excess return of the asset and Market Risk proxied by the Beta[17] instead to acknowledge the fact that only diversifiable risk is priced by the market (see for e.g. Goldreyer (1999); Kreander

[13] The Excess return of an asset is common terminology in Financial Economics and can be defined as the difference between the asset return and the risk-free rate.
[14] The Alpha obtained from CAPM, Fama-French (1993) or Carhart (1997) is typically called Jensen's Alpha, 3-factor alpha and 4-factor alpha, respectively. Throughout this thesis we will only use these specific terms if it's relevant, opting for "alpha" in most cases.
[15] These three results are indicated, respectively, by a positive and statistically significant alpha, negative and statistically significant alpha and non-statistically significant alpha. Levels of significant typically are1%, 5% and 10%.
[16] Tracking error refers to the standard deviation of excess returns of the asset/portfolio.
[17] In this thesis, "Beta" refers to the Market Beta from CAPM while "betas" is a general term referring to any risk-exposure, including exposure to the market but also to other risk-factors such as size, value and momentum from Fama-French based models (namely Fama-French (1993) and Carhart (1997)).

et al. (2005)). However other measures are also used such as the Information, Smith, Tito ratios, among others[18].

The third option consists of "return regressions" because it implies regressing some form of financial return on a set of explanatory variables[19]. This approach differs from the first one in two aspects. First, the dependent variable is not necessarily excess return. Second, explanatory variables don't include benchmark portfolios intended to reproduce risk factors. This approach is usually undertaken concurrently with the other two as most authors choose as a proxy for financial return a risk-adjusted performance measure such as the Jensen's alpha, 3-factor alpha or the 4-factor alpha[20] (Barnett and Solomon, 2006; Capelle-Blancard and Monjon, 2014; Derwall and Koedijk, 2009; Kreander et al., 2005; Lins et al., 2016; Renneboog et al., 2008) or other performance measures as Sharpe and Treynor Ratios, among others (Lee and Faff, 2009). Nonetheless, some authors opted instead for other performance measures, such as absolute returns (Benson et al., 2013; Brammer et al., 2009; Guerard, 1997; Russo et al., 2016) or excess returns (Galema et al., 2008; Nofsinger and Varma, 2014).

The explanatory variables are essentially fund attributes (ex.: size, age, expenses, turnover) and screening characteristics (ex: intensity[21] and/or type of screenings) at the fund level. At the stock level, variables include fundamentals and SRI scores, among others[22]. Some have opted for adding the Beta from CAPM as well (Galema et al., 2008) but other authors have refrained from such approach due to possible error-in-variables problems (Brammer et al., 2009). Some authors opted for combining this approach with the first one by regressing excess returns on not only on the excess return of the market[23] but also variables such as

[18] For more details, see tables in Appendix 1. We don't find relevant to provide extensive lists and definitions of each measure as they are very well known and used in the literature.
[19] Some authors such as Galema et al. (2008) refer to these equations as Fama-MacBeth (1973) regressions because they estimate these regression based on the approach followed by these authors. However, since other authors don't make explicit that they use such method, we prefer the term "return regressions".
[20] As it was previously mentioned, this terminology refers to alphas obtained from CAPM, Fama-French (1993) and Carhart (1997) models, respectively.
[21] Intensity of screening is defined as the number of screens employed by a SRI Fund and will be further analysed in the sub-section "Performance".
[22] For more details, consult the tables in Appendix 1, column "Modifications/Variables".
[23] Used as a regressor in benchmark models (first approach).

fundamentals and SRI scores[24] (Jin et al., 2006; Kurtz and DiBartolomeo, 1999; Nofsinger and Varma, 2014).

Overall, these different methodologies were not able to reach definite conclusions and other approaches to measure performance also yielded contradictory results. As an example, by applying Dynamic Mean-Variance Analysis, Ito et al. (2012) found positive results but Belghitar et al. (2014) found negative results using the Marginal Conditional Stochastic Dominance Model. Diltz et al. (1995) and Blanchett (2010) found mostly non-statistically significant results using statistical tests.

In terms of countries, most authors have focused on the US and UK. As time went by other authors began to focus their attention on different countries, especially in Europe and Asia, and began including multiple regions. Schröder (2007) was the first author to do this in the case of SRI Indices. The evidence shows multiple findings which clearly point out to the necessity of including several countries in the analysis to avoid sample specific evidence (Bauer et al., 2007). However, there are differences between papers that analyse similar countries, even when controlling for time-scope and methodology. For instance, Hill et al. (2007) analysed a period of 1995-2005 and obtained positive results in Europe and non-statistically significant findings in the United States and Asia. On the other hand, Cortez et al. (2012) by analysing the period of 1996-2008 obtain non -statistically and positive results for Europe and United States, respectively.

A closer analysis to the different time-frames also yields no pattern which is exemplified by authors that analysed the same period and came up with different conclusions. Bauer et al. (2005) and Bauer et al. (2006) didn't find a statistically significant difference[25] between SRI and Conventional Investments between 1990-2001 and 1992-2003. On the other hand, Kurtz and DiBartolomeo (1999) report a statistically significant positive alpha for 1990-1991 and

[24] Used as regressors in return regressions (second approach).

[25] Since most authors have opted for obtaining and comparing alpha using benchmark models, in order to avoid needless repletion, we will often use interchangeably expressions such as “difference”; “alpha”, “results”, “findings” and similar expressions.

Climent and Soriano (2011) report negative findings for 1987-2001.[26] However, some recent papers that focused on SRI Indices seem to obtain negative results when analysing more recent periods from 2000's onwards (Belghitar et al., 2014; Charffedinne, 2016; Kurtz and DiBartolomeo, 2011). Therefore, despite evidence that results are time-sensitive, no definite pattern can be estabilished.

We can also see different patterns by analysing the various papers at each level of analysis, we can. In particular, the myriad of results at the stock level contrast with most literature surveys and meta-analysis on the relationship between Corporate Social Responsibility and Financial Performance, which is reported as positive (Margoli and Walsh, 2001, 2003; Orlitzky et al., 2003). However, by taking a closer analysis at more recent literature, as can be seen in Table 16, we found that this relationship is more complex. Despite some positive findings (Derwall et al., 2005; Edmans 2011; Kempf and Osthoff, 2011) many authors reached non-statistically significant results (Auer et al., 2016; Brammer et al., 2006; Erragragui et al., 2016, Mollet and Ziegler, 2014) and some reached negative conclusions (Brammer et al., 2009; Dravenstott and Chieffe, 2011; Trinks et al., 2015).

Anderson and Dakota (2003) can provide an explanation for such a paradox. By conducting a meta-analysis, the authors concluded that higher levels of Corporate Social Performance (hereafter CSP) are more likely to produce positive results but that this relationship depends on how Corporate Financial Performance (CFP) is measured. Specifically, this relationship appears to be stronger when CFP is measured by accounting based measures (such as Return on Assets (ROA)) than when market-based indicators are used (such as the stock market returns), which is the case of most SRI empirical papers.

Some authors have shown that the choice of a benchmark deals a great influence on findings regarding performance (Grinblatt and Titman, 1984). Focusing on Japan, Jin et al. (2006) found that superior performances obtained using TOPIX as a benchmark vanished when using other benchmarks. Ortas et al. (2012) found that while BSCI[27] didn't show any

[26] However, there is evidence of time-varying performance. Given the complexity of the subject, it is explained in more detail in the corresponding section within this Review of the Literature.
[27] Brazil Corporate Sustainability Index.

significant difference in relation to Bovespa and Brasil 50 Index, it underperformed the Brazil Index and BMLC[28]. Ooi and Lajbcygier (2013) found that while SRI Funds showed no statistically significant performance when compared to traditional Fama-French factors, some funds started showing some positive results when the benchmarks were filtered of sin sectors[29]. A recent literature survey by Wallis and Klein (2015) reached similar conclusions. Thus, it is very important to choose carefully choose benchmarks when analysing SRIN to avoid erroneous results.

Some authors claim that the non-statistically significant results of many authors in the literature may be due to the aggregation of different SRI dimensions which may have different and contradictory effects on performance that nullify each other (Derwall et al., 2005; Derwall et al., 2011; Galema et al. 2008). Dimensions are individual criteria such as Environment, Social and Governance and are typically defined and evaluated by different rating agencies such as KLD, Vigeo and Sustainalytics.[30]

Since this evaluation is often carried at the company/stock level, papers at this level can provide a more detailed analysis. Most authors report non-statistically significant results. Some criteria show a positive and statistically significant alpha but they vary between papers (Diltz, 1995; Kempf and Osthoff, 2007; Galema et al., 2008). Results range from good scores regarding Military Involvement and Nuclear Energy (Diltz, 1995) and Community and Employee Relationships (Galema et al., 2008, Kempf and Osthoff, 2007). Other don't find any statistically significant result at all (Brammer, 2006; Vermeir et al., 2005).

Moreover, some authors obtained contradictory results within their work. Galema et al. (2008) reported that only employee relations had a positive effect on performance by using return regressions. However, by using the Carhart (1997) model, the authors found that only community had a statistically significant positive alpha. Erragragui and Revelli (2016) constructed difference portfolios based on two strategies: initiatives engagement and

[28] Brazil Medium-Large CAP Index.
[29] Sin sectors are industries that carry activities not deemed socially responsible such as Tabaco, Gambling, nuclear weapons and others (see for e.g. Trinks et al., 2015).
[30] Examples of the most analysed dimensions and their description are summarised in Table 20.

controversies disengagement[31]. In both cases, most dimensions were shown to be non-statistically significant. However, in the first case a positive and statistically significant alpha was found to associated with governance but, in the second, not only this alpha ceased to be significant and positive, but other two dimensions start showing a negative and statistically significant alpha, human rights and community. These contractions highlight the sensitiveness of results to different approaches, despite not revealing a clear pattern.

At the fund level, authors typically compare the description provided by the Database they used (e.g.: Morningstar, Bloomberg; Lipper, Social Investment Forum) with the fund prospectuses to determine to what specific category SRI Funds belong to. Most authors compare Environmentally friendly funds with traditional SRI Funds since many consider the two types of investment as different categories (see for e.g. Climent and Soriano (2011), Lesser et al. (2014), Silva and Cortez (2016).

However, findings are also very heterogeneous. Many report that Environmental funds clearly surpass traditional SRI and Conventional Funds (Amenc et al., 2010; Ito et al., 2012). Others find that this depends on the analysed subperiod as they tend to underperform in non-crisis periods and show a similar or better performance than SRI and conventional funds during crisis (Lesser et al., 2014; Lesser et al., 2016; Nofsinger and Varma, 2014; Silva and Cortez, 2016). Climent and Soriano (2011) showed that both SRI and Environmental funds significantly underperform during 1987-2001 but are able to match the performance of Conventional funds during 2002-2009. However, Muñoz et al. (2014) found that green funds matched the performance of SRI and conventional funds during any market state.

More recent papers have expanded their scope by analysing funds that screen for other Dimensions besides Environment but no clear conclusion was achieved (Lesser et al., 2016; Barnett and Solomon, 2006; Nofsinger and Varma, 2014; Russo et al., 2016).

[31] The first strategy (initiatives engagement) consists of investing in companies which are proactive in some dimension and the resulting portfolio is the difference between a responsible and a neutral portfolio. The second strategy (controversies disengagement) consists of avoiding companies involved in controversies in certain sectors. The portfolio is the difference between a neutral and irresponsible portfolio.

At the index level, few authors analyse indices related to specific dimensions. Ortas et al. (2013) analysed the Dow Jones Sustainability Index from Asia Pacific (DJSI-AP). By applying Best-of-class screening, the index tends to include companies that strive for improving their clean production practises and technologies. However, no statistically significant difference was found between DJSI-AP and its conventional benchmark. Lesser et al. (2014) analysed SRI and green indices with a global scope. The authors concluded that SRI indices matched the performance of conventional benchmarks during the whole period. On the other hand, Green indices surpass both SRI and Conventional Funds during 2003-2007 but underperform during 2008-2012.

The fact that analysis of dimensions has not provided so far consensual patterns could be the also the result of the difficulty in distinguishing the concept of each dimension separately. The high correlation coefficients between different dimensions reported by some of the mentioned authors give support to this claim (Auer et al., 2011; Galema et al., 2008). A recent paper of Escrig-Olmedo et al. (2010) shows that the lack of transparency and standardization between different rating agencies may also be contributing to the problem. Through an extensive overview of the differences between the different criteria used by rating agencies and Sustainability Indices, the authors concluded that agencies don’t provide extensive explanations of the criteria they use. Moreover, there are differences in the evaluation methodology of companies as different weights are given to the same criteria and each agency and index have a different scoring system.

Other authors have also stressed the ambiguity of certain criteria. According to Kempf and Osthoff the Governance criteria provided by KLD just resulted from the renaming in 2002 of the category "other". This raises questions about the meaning of scores of this dimension. Climent and Soriano (2011) argues that the definition of “green funds” provided by Morningstar is too vague. Lesser et al. (2014) argues that environment should be treated differently from green concerns covered by ESG criteria. Ortas et al. (2013) claimed that clean production systems may not be considered by SRI investments that apply exclusion criteria, even if they have environmental concerns.

We summarise the various definitions of some of the most analysed dimensions in Table 20 and found some examples of such problems. While some agencies define in detail what issues are evaluated within a given dimension, others are very vague about its contents. For instance, EIRIS defined "community" just as community responsiveness. The various descriptions of some very similar dimensions overlap but can contain potentially important differences. For instance, it's not clear whether the concept of "Eco-efficiency" provided by Innovest is consistent with the various "Environment" definitions.

Moreover, at the stock level, all authors base themselves on a single source of SRI sources. However, there are multiple inconsistencies between works that resort to the same agency. The most notorious example can be seen by analysing papers at the stock level which obtain ESG data from KLD, even when analysing similar periods such as the Financial Crisis of 2007. For instance, by analysing the period of 2008-2009, Lins et al. (2016) concluded that high-CSR tend to outperform lower-CSR companies. However, Erragragui and Revelli (2016) found different patterns that varied according to individual dimensions. This highlights the importance of obtaining data from different agencies to obtain a more representative and consistent depiction of SRI.

The great degree of heterogeneity of findings in the literature may reflect that the topic of SRI is still very complex. This can be seen by papers that have analysed the topic of diversification associated with these investments. Theoretically, SRI should underperform conventional investments in terms of risk-adjusted performance because of the application of screening criteria, which restricts the scope of investment and therefore causes a higher degree of risk. (Bauer et al., 2005). However, empirical evidence doesn't support this claim. First, some authors such as Bello (2005) showed that SRI Funds had a similar degree of diversification when compared to Conventional Funds. Becchetti et al. (2015) found that SRI Funds with a more restricted scope performed very similarly to Global SRI Funds[32] and other authors obtained similar results (for e.g., Leite et al. (2014); Muñoz et al., (2014)). The only

[32] The scope of these Funds is mostly the Developed World.

exception was Bauer et al. (2005) who found differences between domestic and international funds located in U.S. and U.K. in terms of performance but no pattern was noticeable.

Other authors have provided additional evidence to the complex relationship between the loss of diversification imposed by the screening processes demanded by SRI and financial performance, by studying the effect of intensity of screening on financial returns (Barnett and Salomon, 2006; Capelle-Blancard and Monjon, 2014; Lee et al., 2010). Intensity of screening consists of the number of screens employed by a SRI fund. It is inversely related to its degree of diversification because higher values imply that more screens are employed and fewer companies are selected. By making return regressions and including intensity of screening as one of the regressors, this variable was shown to have a statistically significant negative impact on performance measures such as the 4-factor alpha, Sharpe Ratio and other performance measures.

However, by including a quadratic intensity of screening, Barnett and Salomon (2006) and Capelle-Blancard and Monjon, (2014) showed that the relationship between returns and intensity of screening is not linear but is shaped by an “U” curve[33]. In particular, the results indicate that while the initial inclusion of screenings tends to have a negative impact at low levels of intensity, at higher levels of intensity the addition of screens can actually improve financial return. According to the authors, this can be explained as the result of the interaction of two opposite effects. As intensity of screening grows more constraints are put in place that reduce diversification possibilities. Fund managers, being aware of these increasingly perverse effects, compensate by choosing stocks with less idiosyncratic risk. The result is that only extreme levels of intensity are optimal: either low levels due to minimal impact on diversification or high levels due to the higher quality of the selected companies.

This loss of diversification is mostly caused by negative screening which excludes companies from sectors not deemed socially responsible. This happens because the so called "sin industries"[34] tend to perform better than SRI and even the market as a whole This is

[33] Lee et al. (2010) were the only authors that didn’t find any evidence to support this claim.
[34] Sin industries are industries that carry activities not deemed socially responsible such as Tabaco, Gambling, nuclear weapons, weapons, and others (see for e.g. Trinks et al., 2015).

confirmed by several authors that showed that companies belonging to these sectors tend to perform better than SRI Funds (Borgers et al., 2015; Capelle-Blancard and Monjon, 2014; Trinks et al., 2015). Nonetheless, even broader negative strategies which simply avoid investing in companies involved in controversies, regardless of the sector, imply some financial loss (Erragragui and Revelli, 2016; Leite et al., 2015).

Another cost that might be taken by SRI fund managers is being forced by fund rules to sell or buy particular stocks at an inappropriate time (Becchetti et al., 2015). In order to study this effect, some authors, such as Girard et al. (2007), Schröder (2004) and Becchetti et al. (2015) analysed the market timing skill of SRI Funds managers. This is described as the ability to allocate capital to stocks before they experience a stock market boom and decreasing their weight before a stock market crash. To that purpose, the authors adopted the approach of Treynor and Mazuy (1966) and Bollen and Busse (2001) which consisted of adding a quadratic term of the excess return of the benchmarks. Other authors, such as Kreander et al. (2005) and Leite et al. (2014), opted for following the approach of Henriksson and Merton (1983) and added a dummy variable to distinguish between positive and negative excess returns. Despite the two slightly different approaches, all authors concluded that there was no statistically significant difference in market timing abilities between SRI and Conventional Funds managers.

2.3 SRI Biases

Based on the literature, we can define a bias as a different characteristic between the investment universe of an asset/portfolio and its benchmark[35]. For example, if a fund has a large cap bias it means that the invested companies are larger in relation to its benchmark. In the case of Fama-French based models, a bias is measured by risk-loadings, which are the loadings on each of the risk-factors (Market, size, value and momentum, profitability and investment)[36].

Since Fama-French based models are often used in SRI literature, many authors have detected such biases. However, by making for the first time an extensive survey, we have found very heterogeneous conclusions. With respect to possible size biases, some papers report a statistically significant Small Cap bias (Areal et al., 2013; Cummings, 2000; Guerard, 1997; Luther et al., 1992) while other authors found a Large Cap bias (Benson et al., 2013; Erragragui and Revelli, 2016; Vermeir et al., 2005) and some reach non-statistically significant results (Becchetti et al., 2015; Brammer et al., 2006). Different results are also achieved with respect to the Value factor, with results ranging from Value bias (Areal et al., 2013; Becchetti et al., 2015; Lee and Faff, 2009), Growth bias (Derwall et al., 2005; Galema et al., 2008; Vermeir et al., 2005) or non-statistically significant bias (Bauer et al., 2006; Brammer et al., 2006; Erragragui and Revelli, 2016). According to most authors, SRI vehicles don't have any statistically significant momentum bias relative to conventional investments (see for e.g. Becchetti et al. (2015), Derwall et al. (2005), Mollet and Ziegler (2014), Renneboog et al. (2008)). However, statistically significant results were found by some authors, being either positive findings (Erragragui and Revelli, 2016) or negative ones (Areal et al., 2013; Leite et al., 2015).

In similarity to performance, findings also vary according to multiple factors. In terms of regions, many authors report a small cap effect in European countries and a large cap one in

[35] Despite being regularly used in the literature, we haven't found a formal definition of "bias". Thus, the definition we present is our own.
[36] Since most biases are detected using these models, "bias" and "risk-loading" will be often used interchangeably. However, one must note that they are different concepts because biases can be detected using other approaches.

the United States (see, among other, Bauer et al., 2005; Becchetti et al., 2015; Hooi et al., 2015; Schröder, 2004). However, some report different findings. Mollet and Ziegler (2014) reports a large cap bias for Europe and non-statistically significant estimates for the US. Cortez et al. (2012) and Silva and Cortez (2016) confirm the small cap bias for Europe but also find it in the US. A small cap in the US was also found by Climent and Soriano (2011) and Areal et al. (2013). Some countries such as Australia don't show any bias (Jones et al., 2008; Bauer et al., 2006).

The small cap bias in Europe may be mostly driven by few countries, namely the UK and Germany (Bauer et al., 2005; Renneboog et al., 2008, Schröder, 2004) but also by Netherlands (Scholtens, 2005) and Sweden (Schröder, 2004). Such conclusions are backed by authors that have focused on other individual European Countries and obtained non-statistically significant results (see e.g.: Leite et al. (2015)).

Some authors regress variables related to the Fama-French factors in order to see the direct impact of screening criteria on possible biases but results are also contradictory. Galema et al. (2008) performed regressions in which the explained variables were Book-to-Market ratios[37] and the regressors were a set of variables that included scores of companies on individual dimensions[38]. They found statistically significant negative estimates for all criteria. This implies that pursuing any particular dimension would cause a portfolio to have a growth bias. Renneboog et al. (2008) opted for a different approach and obtained different results. The authors made individual regressions of each risk loading obtained using the Carhart (1997) model. They concluded that the risk loadings were not affected by any particular dimension but rather by other fund characteristics such as age, size and management fees[39].

---

[37] Book-to-Market Ratios are related to the Fama-French value factor, HML. This connection exists because this factor is reproduced by a portfolio with a long position in High Book-to-Market stocks and a short position in low Book-to-Market stocks.

[38] For a list of dimensions, we refer to Table 20.

[39] First, the authors ran the Carhart (1997) model, obtaining the estimates for each fund for the Market (Beta) and also size, value and momentum risk-exposures. Having obtained those estimates, the authors ran individual regressions. In each one, the regressand was one of the mentioned estimates for each fund and the regressors were a set of variables, namely fundamentals, fund characteristics, screening characteristics, among others.

Some authors using methods very similar in spirit to Fama-French based models have also detected similar biases, despite heterogenous findings. For instance, by adding a size index to the CAPM, Gregory et al. (1997) detected a small cap bias. However, a subsequent work of Kreander et al. (2007) that applied the same approach didn't report any statistically significant bias. Luther et al. (1992) detected a strong small cap bias in UK by comparing the size distributions of ethical unit trusts with conventional ones.

As can be seen in tables from Appendix 1, many authors also report mixed findings about biases by looking at different SRI dimensions. At the stock level, very different findings are reported relative to the size factor. Brammer et al. (2006) and Derwall et al. (2005) obtained non-statistically significant estimates. However, after performing industry-adjusted regressions, the latter detected a large cap bias. Erragragui and Revelli (2016) found a small cap bias associated with a disengagement strategy that focused on community and human rights dimensions and a large cap one related to an initiatives engagement strategy with a focus on governance[40]. Kempf and Osthoff (2007) detected a small cap effect on an aggregate dimension but very different findings when analysing each one individually. In particular, Environment and product dimensions were associated with a small cap bias while community, diversity and negative screenings resulted in a large cap bias. Only employee relations and human rights dimensions showed non-statistically significant estimates.

In relation to the value factor, most authors report a growth bias (Derwall et al., 2005; Galema et al., 2008; Erragragui and Revelli, 2016). Finally, in relation to the momentum factor, most authors were not able to detect a statistically significant momentum bias for any dimension (Brammer et al., 2006; Derwall et al., 2005; Erragragui and Revelli, 2016). Galema et al. (2008) also reported non-statistically significant estimates for most dimensions except for employee relations which showed a positive sign. The only exception found was Kempf and

[40] The initiaves engament strategy consists in investing in companies which are proactive in some dimension and the resulting portfolio is the difference between a responsible and a neutral portfolio. The controversies disengagement strategy consists in avoiding companies involved in controversies in certain sectors. The portfolio is the difference between a neutral and irresponsible portfolio.

Osthoff (2007) which unlike previous authors report very different results for each dimension.

At the fund level, individual dimensions are less analysed and, we have mentioned in the previous subsection “performance”, most authors that follow that course concentrate on Environmental funds. These Funds are shown to have a small cap bias by some authors (Amenc et al., 2010; Climent and Soriano, 2011; Silva and Cortez, 2016) but there's no pattern relative to other biases[41].

The approach with indices is very different as few authors have applied Fama-French based models. One argument put forward by some authors is that that Fama-French factors don't represent the primary assets that constitute SRI Indices but rather a standard investment strategy (Schröder, 2004). Schröder (2007) argues that it's not necessary to apply such models because of three reasons. First, SRI Indices adjust very few times during the year. Second, their behaviour can be well replicated by the respective benchmarks. And third, indices don't follow official styles. Other authors have also referred to these arguments to justify using other models (Ortas et al., 2012; Ortas et al., 2013).

However, the evidence provided by the few authors that have done so shows the importance of its application. Some evidence is provided by Schröder (2007) himself. First, the author reported a statistically significant small cap and value bias for NAI[42] and a growth bias for DSI. Second, the author showed that some alphas obtained from the CAPM model were statistically significant but ceased to be when a Fama-French based model was used instead, which stresses the importance of applying these models to avoid biased results.

Other authors that have applied this model to indices also report statistically significant results, namely Large Cap biases, such as in the case of indices from the DJSI (Lee and Faff, 2009; Scholtens, 2007; Vermeir et al., 2005) and FTSE4GOOD series (Scholtens, 2007; Vermeir et al., 2005). The Domini Social 400 also shares this pattern (see for e.g. Blanchett (2010) and Statman (2006)) but some authors have found non-statistically significant

[41] No such analysis was found at the index level for reasons discussed further ahead in more detail.
[42] Naturaktienindex is a German Index with a global scope.

estimates (Schröder, 2007; Vermeir et al., 2005). Relative to the value factor, few authors have obtained non-statistically significant estimates as well, with most studies reporting a growth bias (Kurtz and DiBartolomeo, 1999; Kurtz and DiBartolomeo, 2011; Vermeir, 2005; Statman, 2006; Scholtens, 2007). The momentum factor is much less studied and most authors don't find statistically significant estimates (Lesser et al., 2014; Blanchett, 2014). Lesser et al. (2014) is one of the few exceptions that didn't find any statistically significant results for any bias estimates. However, the authors also followed a different approach and, instead of analysing each index individually, performed a global analysis by constructing a portfolio comprising many SRI Indices, which could have ignored potential individual differences between the Indices.

Biases were detected in SRI indices also by using other approaches. By analysing the size of the companies belonging to DJSI, Consolandi et al. (2009) found that this index was invested in larger companies than its benchmark. In a similar way, Belghitar et al. (2014) found that 5 of 6 Indexes of the FTSE4GOOD Series reported statistically significant differences in size relative to their respective benchmarks but their nature varied according to the analysed country. More specifically, a large cap was reported in Europe and UK and a small one in US and Global Indices.

Few papers in SRI literature have attempted to provide theoretical theories for what biases may be implied by this kind of investment. Dam and Scholtens (2016) represent an exception. It predicts that more socially responsible firms have higher market to book and return on assets ratios than companies with lower social responsibility standards. This implies that it is theoretically expected that SRI vehicles have a growth and profitability bias in relation to conventional investment vehicles when applying models like Fama-French (2015). The justification for such results is driven by the fact while the former maximizes firm value the latter maximize profits and don’t consider social damage per output. However, to the best of our knowledge, such theories were not yet attested by empirical literature.

Other biases were detected in the literature as well. Recent works in the literature have also applied for the first time the q-theory Model of Hou et al. (2015) to the analysis of SRI. This model adds profit and investment related variables to Fama-French (1993), building on the notion that these variables have a major influence on determining excess returns (Lesser et al., 2014; Lesser et al., 2016). Lesser et al. (2014) found that while green indices have a bias towards companies with low profits and high levels of investment, SRI Indices don't show any statistically significant estimates regarding these variables. These results contrast with other recent authors that have also studied the connection between profitability and CSR during times of market turmoil. In particular, Lins et al. (2016) who studied SRI at the stock level instead, by using return regressions determined that high CSR companies were more profitable during the Global crisis period than low CSR companies. These contradictory results and the fact that this model has rarely been applied in SRI literature justifies the necessity of further analysis. This could be justified by the agglutination of different characteristics of SRI Indices that may cancel each other out, resulting in non-statistically significant results.

Kurtz and DiBartolomeo (1999) argue that differences in return between SRI and conventional investments are not explained by the socially responsible behaviour of the included companies but rather by the sector biases caused by screening processes. Some authors confirmed these tendencies by analysing the exposures to specific industries, namely either by including the returns of selected industries in their regressions (Benson et al., 2013) or a dummy variable to distinguish between different sectors (Lins et al., 2016; Jin et al., 2006). It was found that SRI Funds had a tilt towards information technology, utilities and consumer goods sectors.

Others opted for a different approach and adjusted their regression to possible industry biases. The main approach is the inclusion of industry factors by conducting a Principal Components Analysis (Derwall et al., 2005; Erragragui and Revelli, 2016; Humphrey et al. 2012). Despite not having a straightforward interpretation, as pointed out by the authors themselves, it was found out that the inclusion of industry factors was statistically significant, providing evidence to the existence of industry effects. However, their inclusion did not affect and even intensified the Fama-French original factor exposures. The only exception was found in Humphrey et al. (2012). The standard estimation of the Carhart model (1997) resulted in a statistically significant large and momentum bias. However, after the inclusion of an idiosyncratic risk-mimicking portfolio and industry components, these biases ceased to be statistically significant.

Other authors have opted for industry-adjusted analysis using other approaches but with similar results. Edmans (2011) obtained similar factor exposures by regressing excess return of portfolios on either the risk-free rate or an industry-matched portfolio. Ooi and Lajbcygier (2013) estimated a new Fama-French model by producing new factors that excluded industries deemed not-socially responsible, such as defence, weapons, gambling, among others. The authors confirmed the previous findings by realizing that the factor exposures of the new Fama-French factors were similar to the original ones.

In the case of Indices, Fowler and Hope (2007) point out that SRI Indices are expected to have different industry exposures in relation to their benchmarks. This is mainly due to the application of screening criteria that exclude the inclusion of companies belonging to sectors not deemed socially responsible. However, most authors have refrained from either detecting such biases or conducting industry-adjusted analyses. The only exception found in the literature was Statman (2006). By considering SRI and conventional Industry-weighted indices, the author found that the DSI400 weighted more on the telecommunication and information technology sectors and less on the energy ones relative to S&P500.

Finally, by including a local factor in regression models, a significant home bias was found in SRI Funds of various countries ranging from Australia (Bauer et al., 2006), US (Cortez et al., 2012) and European Countries (Cortez et al., 2012; Leite et al., 2014a; Leite et al., 2014; Leite et al., 2015). This is justified by some authors by the difficulty that SRI Funds managers experience in selecting companies that conform to the socially responsible criteria they must comply with. Moreover, by opting for a more familiar and domestic environment they incur in less monitoring costs (Becchetti et al., 2015).

2.4 Time-varying performance and risk-exposures

Many recent studies have acknowledged the time-varying nature of risk-exposures of SRI vehicles. This prompted many authors to use conditional approaches such as the one from Ferson and Schadt (1996) which consists in allowing betas to be time-variant by including a vector of publicly and pre-determined information variables such as interest rates, exchange rates, dividend yields, among others. All reported to reject the hypothesis of betas to be constant and confirmed the results of unconditional models regarding alpha, which predominantly is reported as non-statistically significant (Bauer et al., 2006; Leite et al., 2014; Renneboog et al., 2008; Schröder, 2004).

Other authors tried to determine if alpha was time-variant as well. To that purpose, some authors adopted the approach of Christopherson et al. (1998) which expands the early method by letting both alpha and betas to be conditional on public information. Findings confirm the time-varying nature of risk factor loadings but there was little evidence of alpha sharing this behaviour (Cortez et al, 2009; Cortez et al., 2012; Leite et al., 2014; Silva and Cortez, 2016).

The adoption of conditional approaches also allows dealing with common econometric problems in return analysis, such as non-normality of residuals which was found in many recent papers (Cortez et al., 2009; Ortas et al., 2012; Ortas et al. 2013; Silva and Cortez, 2016). However, this comes at the expense of degrees of freedom, which could lead to over-parameterization (Busse et al., 2010; Lesser et al., 2014). Some authors such as Galema et al. (2008) dealt with the non-normality problem without using a conditional approach, opting instead for estimating a system of GMM equations, which is simpler and may be extended to other levels of analysis, such as SRI Indices.

Bauer et al. (2006) followed a different approach and performed rolling regressions to test for the stability of the coefficients and found similar results. In particular, SRI Funds underwent a learning phase as alpha increased during the period and a drastic change in investment style during 1992-2003. Specifically, while initially they all had positive exposures to all risk factors, by the end of the period these differences became non-statistically significant.

A simpler approach that also acknowledges the time-varying nature of performance and risk-exposures consists in comparing crisis with non-crisis periods. For the identification of such periods, different approaches are used which are summarised in Table 19. As can be seen, some authors identify crisis periods based on specific criteria such as the definition of Central Banks or specific approaches[43]. Other authors such as Ortas et al. (2013) don't provide any justification for the chosen crisis periods. However, generally speaking, these different approaches don't result in substantially different periods.

One exception occurred with the more recent Sovereigt Debt Crisis which Leite et al. (2015) identified as being from May 2011 and May 2012 while Muñoz et al. (2014) determined to be from October 2009 until January 2013. The great difference is explained by the different approaches. The former based themselves on the approach of Pagan and Sossounov (2003), which considers that crises occur in time periods in which stock prices suffered a slump of at least 20% between a peak and trough. The latter opted for analysing the yields of European sovereign bonds and realized that they began to suffer from the effects of the Sovereign Debt crisis as early as October 2009 and that effect lasted until January 2013.

Other authors opted for identifying different market regimes using markov conditional specifications (Areal et al., 2013; Managi, 2012). Managi (2012) used the Markov Switching model (MS) proposed by Hamilton (1989). This model is intended to replicate processes which are influenced by an unobservable random variable called state variable and was used by Hamilton (1989) to describe US Business cycles. Areal et al. (2013) followed a similar approach and applied a Markov switching specification to a CAPM and Fama-French (1993) model. This approach also avoids similar problems related to the time-varying nature of risk and non-normality returns. However, unlike Conditional Approaches, by depending on just the data and not relying on exogenous variables it also avoids data-mining issues (Areal et al., 2013).

Overall, the main conclusion is that that SRI Funds perform better during crisis than in non-crisis periods (Leite et al., 2015; Lesser et al., 2016, Lins et al., 2016; Muñoz et al., 2014;

[43] As can be seen in table Table 19, these approaches are either based on other papers or devised by the authors themselves.

Nofsinger and Varma, 2014; Soler-Domínguez, 2015; Wu et al., 2017). Some authors opted for comparing SRI with VICE funds in order to intensify the differences between socially and non-socially responsible investments and they also confirmed such findings. While VICE Funds tend to surpass SRI Funds in non-crisis periods, the latter gain the upper hand in times of great turmoil (Areal et al., 2013, Nofsinger and Varma, 2014; Soler-Domínguez et al., 2015). Others choose time periods that were not limited to crisis periods but included them and noticed the absence of negative results (Auer et al., 2016, Climent and Soriano, 2011). Managi (2012) is one of the few exceptions by not finding any statistically significant difference in return means and volatility between SRI and conventional investments across different market regimes.

A possible explanation for these findings is that the investment in social capital allows firms to keep the trust of their stakeholders and investors while overall confidence in companies declines (Lins et al., 2016). The relative positive results of SRF during crises could also come as a compensation for a higher risk associated with an ethical benchmark that is exacerbated during market downturns or booms (Becchetti et al., 2015).

The course of other risk-exposures during different market periods has been less analysed and there's little consistency in the various analysis made in the literature to allow to draw clear conclusions. One apparent exception is the exposure to the Value Factor which seems to increase during periods of market turmoil (Becchetti et al. 2015; Leite et al., 2015; Scholtens, 2005). These results indicate that value companies are more able to keep ethical standards during periods of market turmoil so that they remain included in SRI Indices while growth companies have to choose survival at the expense of environmental, social and governance standards, being more excluded as a result.

As can be seen in Table 19, most mentioned papers have focused on the Global Crisis of 2007 but other market bear periods are also considered such as the recent European Sovereign Debt Crisis of 2011 (Ortas et al., 2013) or the Dot Com Crisis of the early 2000's (Becchetti et al., 2015). Authors that include several crises in their time periods treat them as homogenous and only differentiate them from non-crisis periods by adding dummy variables (see for e.g. Nofsinger and Varma (2014)). However, a recent work of Becchetti et al. (2015)

showed the importance of treating them differently. Despite confirming the findings of most authors by showing that SRI Funds had a better performance in the recent global crisis, they also showed that the opposite occured in the dot com crisis of the early 2000's, presumably due to their higher exposure to high-tech stocks. Thus, such problems could be avoided by conducting an industry-adjusted analysis but so far, to the best of our knowledge, to author has analysed crisis periods within an industry-adjusted analysis framework.

As can be seen by the tables of Appendix 1, most of these approaches are not applied to the analysis of SRI indices. In the case of conditional methods, this is justified by the fact that the inclusion of public information in not expected to affect betas (Schröder, 2004). The only authors that have followed that course were Lesser et al. (2014) who applied the method of Ferson and Schadt (1996) to the study of a portfolio of Green and SRI Indices. Its application was of small relevance, causing only a slight change on the significance level of some negative alphas in relation to the unconditional approach.

However, some authors have applied other similar methods to SRI Indices and obtained interesting results. Ortas et al. (2012) and Ortas et al. (2013) opted for applying the state-space market. This approach consists in letting both alpha and beta from CAPM follow an Autoregressive process[44]. Their results support the claim that alpha and Beta are time-variant. Other authors opted for dividing the time-period into several subperiods and have also confirmed the time-varying nature of the performance of SRI indices (Kurtz and DiBartolomeo, 1999; Kurtz and DiBartolomeo, 2011).

[44] Specifically, the authors allowed alpha and Beta to follow a random walk and a general first order Autoregressive process, respectively.

2.5 Summary

The literature on SRI can be divided into three asset levels of analysis: Funds, Indices and stocks. However, more than half of the literature belongs to the first category. This can pose a problem as their suitability to act as a proxy for SRI is often questioned. Many authors have found evidence that SRI Funds are not transparent, fail to meet their ethical obligations and are increasingly less distinct from Conventional Investments. Moreover, the study of SRI Funds can be distorted by factors such as management fees and transaction costs. Papers at other levels circumvent this problem but Indices remain the best option by representing a real investment option and having a transparent construction process.

In terms of performance, most authors have concluded that there is no statistically significant difference between SRI Vehicles and conventional funds. However, a more detailed analysis shows that there is a lack of consistency in the literature, due to different approaches, resulting in multiple patterns. Most authors use Benchmark models but there are multiple patterns both between different works and within individual papers. They differ extensively in many aspects, namely in the choice of time periods and geographic scope, choice of benchmark, among others.

Most authors have tried to reach more accurate results by analysing individual dimensions. However, this approach also resulted in different patterns, which could be explained by reasons such as the lack of standardization in their definition and evaluation between different SRI Agencies. Moreover, there are inconsistencies between papers that resort to the same SRI Agency, even when controlling for other factors.

This lack of consistency creates difficulties in comparing findings as differences between papers can be attributed to a combination of multiple factors. However, there are some conclusions that can be drawn in relation to certain topics. In particular, there’s no evidence that SRI imposes a restriction on diversification in relation to conventional investments. This can be seen both by analysing the intensity of screening and by comparing Global and Regional SRI Vehicles, which show the same degree of diversification. Such costs are, however, more associated with negative screening by excluding investments in profitable sectors. There’s also no evidence relating that SRI Fund managers have differences in market timing skill in relation to conventional fund managers.

The application of Fama-French (1993) based models by many authors allows analysing findings about biases of SRI in relation to size, value and momentum factors. Nonetheless, in similarity to performance, findings vary extensively according to different factors such as geographic scope, time-period, methodology and dimensions. However, some patterns could be detected in some cases. A large cap is associated with the United States while a small one with Europe, despite the latter being most likely present in just a few countries such as Germany and the UK. By looking at individual dimensions, at the stock level most authors have found a value bias while authors who focused on Environmental Funds have found a small cap bias.

The approach with SRI Indices is very different as few authors have refrained to apply such methodology. However, the extant literature shows evidence of the importance of its application. In particular, many indices were found to have a large and growth bias, such as indices from the FTSE4GOOD and DJSI Series, even when using other approaches.

Other biases were detected in the literature as well. A few recent works have applied Fama-French models that add profit and investment related variables but findings are contradictory. Many authors have also performed analysis adjusted for the presence of industry biases, which were proven to exist in many cases. However, in similarity to the application of Fama-French models, most of these papers belong to the fund level. Finally, many authors have detected a strong local bias in many SRI Funds.

Many authors have acknowledged the possibility that alpha and risk-exposures are not constant over time. Most authors have applied conditional models and found evidence to support the time-varying nature of Beta and other risk-exposures. However, no such evidence was found for Alpha. Some opted for simpler approaches such as comparing crisis and non-crisis periods. Despite different methods of identification of these periods, overall, it was found that SRI tend to outperform Conventional Investments during times of market turmoil. There’s no clear pattern regarding most risk-exposures, except for the exposition to the value factor, which increases during market downturns. These approaches, however, continue to be rarely applied to the analysis of SRI Indices.

2.6 Contributions to the literature

We conducted a very comprehensive Review of the Literature which allowed to obtain a more accurate perspective of SRI literature as a whole. This constitutes an important contribute to the literature since, to the best of our knowledge, we were also the first to review findings not only related to performance but also to biases uncovered mostly by the application of Fama-French based models.

Based on this knowledge, we contribute in two additional ways. First, we analyse SRI Indices from three different Series, FTSE4GOOD, MSCI ESG and STOXX ESG. The indices from each series belong to four categories according to which region they cover: Global, Europe, Japan and United States. As we have shown, SRI Indices are analysed less frequent in relation to funds, despite providing a more accurate depiction of the SRI Universe. Moreover, to the best of our knowledge, we are the first to conduct an analysis of SRI Indices in this manner. By choosing Indices from different sources and geographic scopes we obtain more accurate and representative results of SRI as a whole.

Our second contribution consists on the novel methodology we apply to the analysis of SRI Indices which is innovative in multiple and important ways:

- We apply not only CAPM but also Fama-French based models, which were rarely applied to the analysis of SRI Indices. By taking into account possible biases, their application allows not only to obtain more accurate results in terms of performance, but also to analyse differences between companies that are included in SRI Indices and those that are excluded in terms of size, value and momentum, profitability and investment profiles.
- We also conduct an industry-adjusted analysis by adding industry portfolios to the mentioned models by a Principal Components Analysis. Such method was never applied to the analysis of SRI Indices and allows to obtain more robust results in the presence of industry effects and also to analyse their impact in extant biases.

- We take into account findings of the literature regarding the possible time-varying nature of performance and risk-exposures by adding a dummy variable to allow a comparison between crisis and non-crisis periods.
- We estimate the models using a Generalized Method of Moments System Framework, which was never applied to the analysis of SRI Funds and allows to deal with the common problem of non-normality of returns.

3. Data and Methodology

3.1 SRI Indices

For our thesis, we selected 7 SRI Indices from the FTSE4GOOD series, 4 indices from MSCI Global Sustainability Indices and 4 indices from STOXX ESG[45] Leaders Indices. They were selected based on selection criteria that we describe and justify in Appendix 2. Information about these indices and their official benchmarks[46] is available at Table 1. The Indices we selected have both a global[47] and regional scope that focus on the United States, Europe and Japan[48]. As we explain in more detail in appendix 2, this selection of SRI Indices allows to obtain more representative and accurate results about SRI by relying on different sources and geographic scopes. Each Index can be divided into two categories: benchmark[49] and tradable. A benchmark index covers the entire universe of the companies that pass the screening criteria of the SRI Index while a tradable index concentrates on the 50 or 100 largest companies[50].

We followed the approach of many authors in the literature that have studied SRI indices and opted for gathering daily returns from Thomson Reuters Datastream Database (hereafter

[45] The complete designation of these indices is STOXX ESG Leaders but we opted to mostly use this term for simplication.
[46] Official benchmarks are Conventional Indices from which the SRI Indices are composed by the application of screening criteria. We confirmed the official benchmark of each SRI Index by looking into freely available information provided by the company associated with each Series.
[47] Global SRI Indices and their respective official conventional benchmarks from the three Series mostly focus on Developed Markets, on countries belonging mostly to Europe, United States and Japan. The same is true for the Fama-French factors with a global scope, as we show in appendix 3.
[48] In the case of STOXX ESG Series, STOXX AP ESG Leaders 50 and its respective benchmark cover mostly Japan despite also covering other countries such as Australia, Singapore and New Zealand. In the case of SRI Indices that cover North America, Canada. However, in both cases the remaining countries have very modest weights.
[49] The term "benchmark" is often used to refer to either the official benchmark of a SRI Index or a SRI Benchmark Index which imposed no restrictions to the size of the companies that comprises. To avoid confusion, we will use the term "official benchmark" to refer to the official benchmark of each SRI Index.
[50] In the case of the STOXX ESG Leaders Series, these Indices are called "Blue-Chip". However, we prefer the term tradable since it's more used in SRI literature (see for e.g. Collison et al. (2008), Sun et al (2011)).

Datastream) in order to obtain as many observations as possible[51]. Returns of index i at day t will be calculated as:

$$r_{i,t} = \ln\left(\frac{p_{i,t}}{p_{i,t-1}}\right) \qquad (3.1)$$

Where $p_{i,t}$ represents the closing price of index i at day t, adjusted by dividends and capital increases on day t and ln is the natural logarithm. In order to avoid biased results due to the fact that indices come from several countries and are priced in different currencies, we procure to obtain returns converted into dollars, as many authors have also chosen this approach[52]. When not possible, we convert their values into dollars by using exchange rates also obtained at Datastream[53].

Time spans vary between Indices belonging to different Series but are rather similar within each one. Since we need to analyse each Series using balanced samples to avoid inconsistencies and due to our methodology, each Series will be analysed in the following time periods: 31/11/2004 – 31/05/2017 (FTSE4GOOD), 01/07/2007 – 31/05/2017 (MSCI ESG), 28/07/2013 – 31/05/2017 (STOXX ESG).

---

[51] See for e.g. Hill et al. (2007), Collinson et al. (2008), Charfeddine (2016), Schröder (2007), among others.
[52] See for e.g. Ortas et al., (2012); Schröder (2007), among others.
[53] To this purpose, in the formula in (3.1) we follow Ortas et al.,(2012) and divide $p_{i,t}$ and $p_{i,t-1}$ by their respective daily exchange rates, also obtained at Datastream.

Table 1 - List of SRI Indices.
Source: Adapted from database Datastream. We provide a list of all the SRI Indices and their respective official benchmarks, identified by their respective Datastream code. Start Date indicates the date of the first common return observation of each SRI index and its respective official benchmark.

SRI Index	Code	Type	Official Benchmark Index	Code	Start Date
FTSE4GOOD EUR	FT4GBEU	Benchmark	FTSE DEV EUR	AWDVER$	01/07/1996
FTSE4GOOD EUR 50	FT4EU50	Tradable	FTSE DEV EUR	AWDVER$	01/07/1996
FTSE4GOOD GLB	FT4GBGL	Benchmark	FTSE DEV	AWDVLP$	01/07/1996
FTSE4GOOD GLB 100	FT4G100	Tradable	FTSE DEV	AWDVLP$	01/07/1996
FTSE4GOOD JAP	FT4GBJP	Benchmark	FTSE JAP	WIJPANL	31/11/04
FTSE4GOOD US 100	FT4U100	Tradable	FTSE US	WIUSAML	01/07/1996
FTSE4GOOD US	FT4GBUS	Benchmark	FTSE US	WIUSAML	01/07/1996
MSCI EUROPE ESG	MSEUSG$	Benchmark	MSCI EUROPE	MSEROP$	28/09/2007
MSCI JAPAN ESG	MSJPSG$	Benchmark	MSCI JAPAN	MSJPANL	28/09/2007
MSCI NORTH AMERICA ESG	MSNMSG$	Benchmark	MSCI NORTH AMERICA	MSNAMR$	28/09/2007
MSCI WORLD ESG	MSWESG$	Benchmark	MSCI WORLD	MSWRLD$	28/09/2007
STOXX AP ESG LDRS 50	SAPEL5$	Tradable	STOXX ASIA/PACIFIC 600	DJSTAP$	24/05/2013
STOXX EU ESG LDRS 50	SEUEL5E	Tradable	STOXX EUROPE 600	DJSTOX$	24/05/2012
STOXX GLOBAL ESG LEADERS	SGESGL$	Benchmark	STOXX GLOBAL 1800	DJS180$	24/05/2012
STOXX N.AMR ESG LDRS	SNAEL5$	Tradable	STOXX NORTH AMERICA 600	DJSTAM$	24/05/2012

3.2 Descriptive Statistics

Table 2 shows descriptive statistics of the daily total returns of all the SRI indices and respective benchmarks, which are paired up for convenience. We follow the approach of many authors in the literature and calculate the Sharpe Ratio which is defined by the following expression:

$$\frac{E(r_i - r_f)}{\sigma_i} \quad (3.2)$$

Where r_i is the return of index i, r_f is the risk-free rate[54] and σ_i is the standard deviation of the excess returns of index i[55]. It's apparent that there are substantial differences between the SRI Indices of the three Series. On the one hand, SRI Indices from the FTSE4GOOD Series appear to underperform their conventional benchmarks. On the other hand, SRI Indices from the other series show mixed behaviours, with no clear pattern even when comparing Indices with similar geographic scope.

As we have shown in the Review of the Literature, many authors have detected that returns from SRI Vehicles and their conventional counterparts are not normally distributed. We conducted a Jarque-Bera Test for all indices to confirm this claim. The Jarque-Bera test evaluates the null hypothesis that a variable follows a Normal Distribution. As can be seen, the null hypothesis is rejected for the returns of all Indices at 1% level of significance.

[54] As it will be further explained, we use the US 1 month treasury bill as a proxy for the risk-free rate for all Indices.

[55] This expression of the Sharpe Ratio is based on the work of Ortas et al. (2013).

Table 2 - Descriptive Statistics of the Daily Excess Returns of the FTSE, MSCI and STOXX Series.
Source: Adapted from econometric software Eviews 9. Sample (observations): FTSE4GOOD - 02/12/2004 – 31/05/2017 (3260); MSCI ESG - 01/10/2007 – 31/05/2017 (2523); STOXX ESG - 28/07/2013 – 31/05/2017 (1048).* significant at 10%; ** significant at 5%; *** significant at 1%.

Index	Mean	Maximum	Minimum	Std. Dev.	Sharpe	Skewness	Kurtosis	Jarque-Bera	Obs.
FTSE4GOOD EUR 50 RF	0.008637	11.34137	-10.0055	1.428424	0.006047	-0.10102	11.43102	9660.864***	3260
FTSE DEV EUR RF	0.01664	10.84125	-10.2673	1.428805	0.011646	-0.1586	10.98213	8668.203***	3260
FTSE4GOOD EUR RF	0.014991	11.08986	-10.0691	1.436089	0.010439	-0.15325	11.0258	8762.257***	3260
FTSE DEV EUR RF	0.01664	10.84125	-10.2673	1.428805	0.011646	-0.1586	10.98213	8668.203***	3260
FTSE4GOOD GLB 100 RF	0.015169	10.17479	-7.26312	1.091228	0.013901	-0.19214	13.03417	13696.37***	3260
FTSE DEV RF	0.02163	9.088975	-7.2091	1.021074	0.021184	-0.48111	12.81236	13204.12***	3260
FTSE4GOOD GLB RF	0.019413	9.338524	-6.89202	1.06876	0.018164	-0.31549	12.12024	11352.53***	3260
FTSE DEV RF	0.02163	9.088975	-7.2091	1.021074	0.021184	-0.48111	12.81236	13204.12***	3260
FTSE4GOOD JAP RF	0.010687	11.59571	-9.61187	1.401308	0.007626	-0.17342	8.196667	3684.565***	3260
FTSE JAP RF	0.011341	11.38374	-9.41068	1.372011	0.008266	-0.19315	8.321503	3866.852***	3260
FTSE4GOOD US 100 RF	0.024613	11.46365	-9.50926	1.194731	0.020601	-0.19574	15.0372	19702.29***	3260
FTSE US RF	0.025866	10.86879	-9.31352	1.182156	0.02188	-0.35067	14.83007	19076.76***	3260
FTSE4GOOD US RF	0.025187	11.44494	-9.42304	1.198162	0.021021	-0.20229	14.83343	19042.97***	3260
FTSE US RF	0.025866	10.86879	-9.31352	1.182156	0.02188	-0.35067	14.83007	19076.76***	3260
MSCI EUR ESG RF	0.005019	10.26111	-9.42767	1.518184	0.003306	-0.15253	9.401086	4317.163***	2523
MSCI EUR RF	0.003451	10.76045	-10.1778	1.541157	0.002239	-0.13032	10.01863	5185.717***	2523
MSCI JAP ESG RF	0.005458	11.6821	-9.55297	1.457676	0.003744	-0.15867	8.369229	3041.196***	2523
MSCI JAP RF	0.002794	13.06188	-10.4351	1.519336	0.001839	-0.33927	10.2495	5573.28***	2523
MSCI NA ESG RF	0.022346	9.697747	-9.58578	1.271403	0.017576	-0.45213	12.50984	9593.158***	2523
MSCI NA RF	0.023795	10.42764	-9.49625	1.286834	0.018491	-0.43018	13.47761	11618.47***	2523

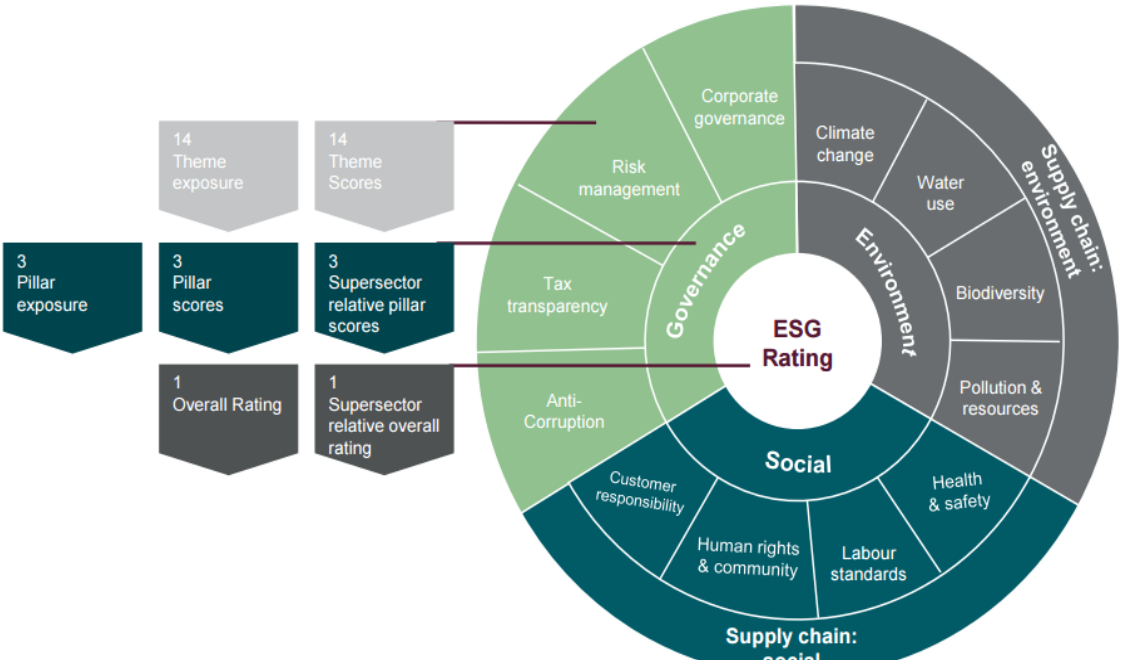
MSCI WOR ESG RF	0.015143	8.56331	-7.19835	1.114426	0.013588	-0.45178	11.13649	7045.36***	2523
MSCI WOR RF	0.015416	9.097467	-7.31687	1.120175	0.013762	-0.46891	11.65727	7971.402***	2523
STOXX AP ESG 50 RF	0.018772	5.074778	-5.02601	1.000503	0.018763	-0.31545	6.646412	597.987***	1048
STOXX AP 600 RF	0.025186	4.281399	-5.3169	0.988618	0.025476	-0.391	5.788992	366.3638***	1048
STOXX EUR ESG 50 RF	0.02331	4.390416	-9.38486	1.015075	0.022964	-0.94528	11.98965	3684.943***	1048
STOXX EUR 600 RF	0.017065	4.534282	-11.1118	1.09686	0.015558	-1.08551	14.50102	5981.751***	1048
STOXX GLB ESG RF	0.033388	2.598726	-5.11454	0.688366	0.048503	-0.85257	8.346336	1375.099***	1048
STOXX GLB 1800 RF	0.032743	3.445505	-7.84925	0.863521	0.037918	-0.98917	11.71311	3486.001***	1048
STOXX NA ESG 50 RF	0.041855	3.659925	-3.99521	0.770308	0.054335	-0.40562	5.714953	350.6038***	1048
STOXX NA 600 RF	0.037423	3.127047	-3.95273	0.768541	0.048694	-0.36742	5.487956	293.8723***	1048

3.3 Brief Description of the Series

In this section, we provide a brief description of the main features and the screening process of the three Series of Indices we selected to analyse in our Thesis[56].

FTSE4GOOD represents the most established series of our analysis, with the first index released on November of 2001. Responsibility standards are provided by the Financial Times Stock exchange company (hereafter FTSE), owned both by the Financial Times and London Stock Exchange. MSCI Global Sustainability Indices were officially launched by Morgan Stanly Capital International (hereafter MSCI) in 2010 but have data as early as 2007. The most recent STOXX ESG Leaders Indices appeared recently in 2011 and are of the responsibility of Bank Sarasi, with Sustainability Data provided by Sustainalytics.

Figure 1 – Example of an ESG Model from FTSE
Source: FTSE



[56] The following information was based on Sun et al. (2011), who made an excellent analysis of the status of SRI Indices, and on websites of the three series: www.ftse.com; www.msci.com; www.stoxx.com.

There are differences between the Series in terms of what specific criteria are evaluated, how they are weighted and how they are scored[57]. However, they are all centred around three pillars: environment, social, governance. Figure 1 represents an example of the ESG model followed by FTSE. MSCI ESG and STOXX ESG adopt similar approaches.

The selection process takes the respective universe indices as a starting point. It typically involves two steps. The first step excludes companies involved in controversial areas such as Alcohol, Tobacco, Gambling, Weapons, Nuclear Power, among others. Companies are also excluded if they breach established International Principles such as the UN Global Compact Compliance Principles.

The second step involves evaluating the remaining companies from the respective universe index. This evaluation is based on models that evaluate the performance of the companies on a variety of indicators related to the three pillars of Socially Responsible Investment we mentioned: environment, social and governance[58]. These systems are oversighted by independent committees comprising elements from the investment community, NGO's, academia, communities and companies. Ratings are only based on public available information.

Finally, companies that are able to achieve a minimum ESG score are selected into an SRI index. Indices are reviewed once or twice a year and companies already belonging to an index also have to achieve a minimum ESG score to continue to be listed in it. However, this threshold is typically lower in relation to companies already inside the index, which implies that less effort is required to be considered socially responsible.

[57] For more details about the specific characteristics of screening process of each Series, see Sun et al. (2011).

[58] As we have shown in the Review of the Literature, there is no consensus in the definition of the three concepts.

3.4 Models

In order to analyse the returns of the three series, we apply several models in our analysis.

(3.3) $$R_{it}\text{-}R_{ft}=\alpha_i+\beta_{1i}(R_{mt}\text{-}R_{ft})+\varepsilon_{i,t}$$

The first model is CAPM and is represented by equation (3.3). R_{it} represents the return of SRI index i at day t, α_i is the Jensen's alpha (Jensen, 1968) of SRI index i, R_{mt} represents the return of benchmark index m at day t, R_f represents the risk-free rate at day t and $\varepsilon_{i,t}$ is the error term. Thus β_{1i} represents the Beta in which values above (lower) than 1 imply that SRI Indices show higher (lower) risk than their official conventional benchmarks. We apply this model in order to determine if the behaviour of the SRI Indices can be replicated by their respective official conventional benchmarks. To this purpose, we also follow the approach of Schröder (2007) by conducting a spanning test, which tests the following hypotheses:

$$H0:\ \alpha = 0\ \wedge\ \beta = 1$$

$$H1:\ \alpha_i \neq\ 0\ \forall\ \beta_i \neq 1$$

A rejection of the null hypothesis implies that the SRI Index cannot be replicated by the respective benchmark index. Such result would imply that the SRI index has a different risk and/or return profile relative to its benchmark and would give further evidence to the necessity of the application of more complex models.

(3.4) $$R_{it}\text{-}R_{ft}=\alpha_i+\beta_{1i}(R_{mt}\text{-}R_{ft})+\beta_{2i}SMB_t+\beta_{3i}HML_t+\varepsilon_{i,t}$$

The second model is Fama-French (1993) which adds two risk-mimicking portfolios to the CAPM model, SMB and HML and is represented by equation (3.5). SMB (Small Minus Big) is a zero-investment[59] portfolio with a long position in small cap stocks and a short position in Large cap stocks and is intended to reproduce the size effect. HML (High minus Low) is a zero-investment portfolio with a long position in High Book-to-Market stocks and a short position in low Book-to-Market stocks and reproduces the

[59] A zero-investment portfolio is a portfolio where the sum of the weights of each asset totals 0. For instance, the SMB portfolio constructed by Fama-French (1993) has the following composition: SMB= 1/3 (Small Value + Small Neutral + Small Growth) – 1/3 (Big Value + Big Neutral + Big Growth). Has can be seen the total sum of the weights of each asset is zero. Other zero-investment portfolios are constructed in a similar way. More details on the construction ans scope of these Portfolios is provided in appendix 3.

Value effect. In case of positive estimates, it is representative of a value bias, which implies that SRI Indices are more invested in companies with higher Book-to-Market Ratios. In case of negative exposures, it reveals exposures to companies with lower Book-to-Market Ratios and hence greater growth prospects.

$$R_{it}\text{-}R_{ft}=\alpha_i+\beta_{1i}(R_{mt}\text{-}R_{ft})+\beta_{2i}SML_t+\beta_{3i}HML_t+\beta_{4i}WML_t+\varepsilon_{i,t} \quad (3.5)$$

The third model is Carhart (1997). This model adds a portfolio intended to capture the momentum anomaly detected by Jegadeesh and Titman (1993). This portfolio is WML (Winners minus Losers) and has a long position in winner stocks and a short position in loser stocks[60].

$$R_{it}\text{-}R_{ft}=\alpha_i+\beta_{1i}(R_{mt}\text{-}R_{ft})+\beta_{2i}SML_t+\beta_{3i}HML_t+\beta_{4i}WML_t+\beta_{5i}RMW_t+\beta_{6i}CMA_t+\varepsilon_{i,t} \quad (3.6)$$

Finally, we also apply the Fama and French (2015) model, represented in the above equation, which adds profit and investment portfolios[61]. The first portfolio is RMW (Robust minus Weak) and consists of a zero-investment portfolio, with a long position in companies with robust profits and short in companies with weak profits. The second portfolio is CMA (Conservative minus Aggressive) and consists of a long position in companies with a conservative investment policy (low investment) and a short position n companies with an aggressive investment strategy (high investment).

The application of such models to the analysis of SRI Indices accomplishes two goals. First, it provides a more accurate estimate of risk-adjusted performance if biases are detected. Second, since the only difference between SRI Indices and their respective official conventional benchmarks is the application of screening criteria, the application of these models also creates the possibility of discovering its impact. Specifically, it allows to determine what attributes are favoured by SRI screening processes and hence what characteristics differentiate companies that are included in SRI Indices from those

[60] Winner (loser) stocks are stocks that achieved positive (negative) returns in the previous 12 months.

[61] In reality, Fama French (2015) model does not have the momentum variable, adding just the profit and investment portfolios to the Fama-French (1993) model. The authors excluded the momentum variable because they found it not to be statistically significant. However, since that model was tested with conventional stocks, that may not be appropriate for the case of SRI. For that reason and for simplification, we keep the momentum variable when adding profit and investment related variables. Thus, whenever we refer to the Fama French (2015) model, we are referring to the model represented in equation (3.7).

that are excluded. For instance, if a small cap bias is detected in the analysis of some SRI Indices, it implies that their screening processes tend to favour the inclusion of smaller companies in relation to the official benchmark. A similar reasoning follows to the other factors included in these models.

In similarity to many authors, we retrieved data for the risk-free rate and returns of the zero-investment portfolios from Kenneth French Library[62]. These portfolios cover different geographical scopes: Global, Global ex-US, Europe, Japan, Asia Pacific ex Japan, North America. This allows us to we use zero-investments portfolios that match the geographical scope of each of the analysed SRI Indices. For that reason, given the geographic scope of the SRI Indices we analyse, we select portfolios with the following scope: Global, Europe, Japan, North America. The proxy for the risk-free rate is the US one-month treasury bill[63]. In appendix 3, we provide a more detailed description of the geographic scope of the Fama-French 5 Factors.

As can be seen, each model we present gradually adds more variables to previous ones, culminating in the Fama-French (2015) model. The gradual application of more complex models allows to determine the impact of the addition of new variables on extant ones and, thus, to determine more accurately if results are robust or sensitive to the application of more complex models.

[62] Available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.
[63] We opted for using a common risk-free rate for all SRI Indices because Fama-French (1993) also used that approach, as can be seen by data available at Kenneth French Library. Moreover, other authors that have analysed SRI Indices with different geographical scopes also opted for using a US risk-free rate such as Schröder (2007), Managi (2012) and Lesser et al. (2014), among others.

3.5 Generalized Method of Moments System Framework

We adopt an approach similar to Galema et al. (2008) and estimate the models using the Generalized Method of Moments (henceforth GMM) estimation procedure[64]. We estimate it in the form of three systems, each using a balanced sample of observations from the respective Series[65]. This method has several advantages in relation to the typical Ordinary Least Square (henceforth OLS) procedure. First, it allows the errors of equations to be contemporaneously correlated. As can be seen in Table 3, this is justified by the fact that many SRI indices share very high coefficients of correlation, especially Global and regional indices that focus on North America. Second, by relying on weaker assumptions, it circumvents the problem of non-normality of returns which has been detected by many authors in the literature and we confirmed, by the normality tests we conducted on the SRI Indices we selected.

In similarity to Galema et al. (2008), we test the null hypothesis that each regressor is null throughout each System. Moreover, since we test each model separately to see the impact of the addition of new variables on the existing ones, we also perform such tests on groups of new regressors introduced by more complex models. Thus, our analysis covers two levels: At the index level, we analyse the statistical significance of a particular regressor in the regression of an index using the traditional individual t-statistics; At the Series level, we analyse the statistical significance of a regressor and group of regressors throughout a particular System of SRI Indices belonging to a Series by recurring to J-statistics.

Finally, we also follow the usual approach and apply the Newey-West (1987) estimation procedure, which provides results robust to heteroskedasticity and serial correlation.

[64] The assumptions of the GMM Procedure are explained in detail in Appendix 2.

[65] We opted for analysing a balanced rather than an unbalanced sample because authors in the SRI literature that have estimated systems of equations also opted for this approach (see, for e.g., Galema et al. (2008), Schröder (2007)). Moreover, Hayashi (2000), who provides a very good description of the GMM procedure, is one among many authors that assumes that n is equal between equations when describing the GMM Procedure for multiple equations.

Table 3 - Correlation between the FTSE4GOOD, MSCI ESG and STOXX ESG Series

Source: Adapted from econometric Software Eviews 9. Shaded areas denote correlations of indices belonging to the same series. The values of the correlations take into account the common sample period of all SRI Indices (28/07/2013 - 31/05/2017). Within each Series, similar values are also obtained when considering the sample period mentioned in section 3.1. FT4G – FTSE4Good.

SRI Index	FTSE4GOOD EUR 50	FTSE4GOOD EUR	FTSE4GOOD GLB 100	FTSE4GOOD GLB	FTSE4GOOD JAP	FTSE4GOOD US 100	FTSE4GOOD US	MSCI EUROPE ESG	MSCI JAPAN ESG	MSCI NORTH AMERICA ESG	MSCI WORLD ESG	STOXX AP ESG LDRS 50	STOXX EU ESG LDRS 50	STOXX GLOBAL ESG LEADERS	STOXX N.AMR ESG LDRS 50
FTSE4GOOD EUR 50	1.000														
FTSE4GOOD EUR	0.992	1.000													
FTSE4GOOD GLB 100	0.851	0.847	1.000												
FTSE4GOOD GLB	0.881	0.887	0.986	1.000											
FTSE4GOOD JAP	0.189	0.184	0.231	0.314	1.000										
FTSE4GOOD US 100	0.530	0.534	0.875	0.824	0.042	1.000									
FTSE4GOOD US	0.536	0.540	0.876	0.828	0.042	0.999	1.000								
MSCI EUROPE ESG	0.980	0.993	0.838	0.879	0.188	0.527	0.533	1.000							
MSCI JAPAN ESG	0.189	0.184	0.228	0.312	0.989	0.039	0.039	0.187	1.000						
MSCI NORTH AMERICA ESG	0.571	0.575	0.871	0.839	0.048	0.965	0.971	0.564	0.044	1.000					
MSCI WORLD ESG	0.807	0.814	0.971	0.980	0.290	0.875	0.881	0.809	0.290	0.915	1.000				
STOXX AP ESG LDRS 50	0.360	0.361	0.395	0.476	0.840	0.152	0.154	0.359	0.852	0.173	0.440	1.000			
STOXX EU ESG LDRS 50	0.977	0.987	0.836	0.875	0.168	0.533	0.539	0.982	0.167	0.569	0.801	0.337	1.000		
STOXX GLOBAL ESG LEADERS	0.938	0.959	0.873	0.925	0.296	0.593	0.601	0.954	0.297	0.646	0.873	0.506	0.949	1.000	
STOXX N.AMR ESG LDRS 50	0.602	0.607	0.859	0.840	0.079	0.904	0.912	0.592	0.076	0.956	0.904	0.222	0.598	0.696	1.000

3.6 Robustness Tests

We complement our analysis with procedures that can help confirm the robustness of our results. Each of these procedures is applied to the Fama-French (2015) model since this is the most complete model we use. The procedures are summarised below and will be explained in more detail in the following sections:

- Industry Adjusted Regressions[66] – Addition of Industry factors to test for Industry-effects.
- Single GMM System– Application of a single GMM system to all equations. This procedure allows making more reliable comparisons between SRI Indices of different Series.
- Regional Portfolios – We construct portfolios of Indices with similar geographical scope to analyse what patterns emerge in different regions.
- Crisis Dummy's – Addition of a dummy to allow alpha and risk-exposures to vary between crisis and non-crisis periods.
- Variance Inflation Factor Analysis – We analyse the Variance Inflation Factors of each individual regression to test for the presence of multicollinearity.

[66] If industry effects are detected, these procedure will combined with the remaining ones in order to obtain industry-adjusted results.

3.6.1 Industry Adjusted Regressions

Since many authors have detected that SRI may be tilted towards certain industries, it's necessary to adjust our models to the possible presence of industry biases. For example, let's suppose that a statistically small cap bias is detected in the original Fama-French (2015) model. In the absence of industry biases, this result would imply that the screening process of SRI indices is more favourable to the inclusion of smaller companies. However, in the presence of industry biases, another possible explanation for such result could be a bias towards industries that happen to have smaller companies than the average industry.

In order to adjust for the presence of industry biases, we adopt a similar procedure to that of other authors such as Erragragui and Revelli (2016), Derwall et al. (2015) and Humphrey et al. (2012). To the best of our knowledge, this approach was never applied to the analysis of SRI Indices. It consists of adding three industry factors to the Fama-French (2015) model, resulting in the following model:

$$(3.7) \quad R_{it}\text{-}R_{ft} = \alpha_i + \beta_{1i}(R_{mt}\text{-}R_{ft}) + \beta_{2i}SML_t + \beta_{3i}HML_t + \beta_{4i}WML_t + \beta_{5i}RMW_t + \beta_{6i}CMA_t + \beta_{7i}IP1_t + \beta_{8i}IP2_t + \beta_{9i}IP3_t + \varepsilon_{i,t}$$

The industry factors (principal components) are represented by IP1, IP2 and IP3. Each industry factor is obtained by conducting a Principal Components Analysis on the residuals of the following OLS equations of the Fama-French (2015) model:

$$(3.8) \quad R_{it}\text{-}R_{ft} = \alpha_i + \beta_{1i}(R_{mt}\text{-}R_{ft}) + \beta_{2i}SML_t + \beta_{3i}HML_t + \beta_{4i}WML_t + \beta_{5i}RMW_t + \beta_{6i}CMA_t + \varepsilon_{i,t}$$

Where R_{it}-R_{ft} represents the difference between the returns of a portfolio composed of stocks belonging to industry i and risk-free rate. Each industry-portfolio represents a portfolio of stocks of one of 49 industries listed in Table 24, whose data we retrieved from Kenneth French Library[67].

[67] The industries are listed in Appendix 4. Since there's no differentiation between geographic scopes relative to these portfolios, we assume they have a global scope and thus, in the Fama-French (2015) model, we use Fama-French zero-investment portfolios with a global scope. Moreover, we lose some observations because these portfolios don't have data for certain dates.

A Principal Components Analysis consists of analysing the structure of the variance-covariance of a set of variables through linear combinations of those variables. These linear combinations are called Principal Components. The first principal component is a linear combination that maximizes the variance of a set of variables. The second principal component maximizes the remaining variance subject to the restriction of being uncorrelated with the first principal component. Subsequent principal components are obtained in a similar way (Johnsson and Wichtern, 2007).

One of the major advantages of such approach is the reduction of Variables (Johnsson and Wichtern, 2007). In the particular case of the analysis we conduct, instead of adding 49 individual industry-adjusted portfolios, this approach enables the addition of just a couple of variables, called principal components, that represent most variation of the excess return of these industry-portfolios.

3.6.2 Single GMM System

The application of a GMM System to each of the three Series allows performing a representative analysis since it considers most observations of each Series. However, it creates some difficulties in comparing similar[68] SRI Indices from different Series. Differences could be justified not only by differences in the screening process but also because of different time-spans. For instance, when estimating a GMM System for each Series one could find different value exposures between FTSE4Good Europe and MSCI ESG Europe that could be caused by the fact that the first index has more observations than the second one.

To overcome such obstacles, we restrict our analysis to the common period of all the three Series. This results in a sample period of 19/9/2012-30/6/2016 with 987 observations for each index and 15612 for the overall System. This procedure assures that any differences found between similar SRI Indices from different Series are just the result of differences between the screening processes[69].

[68] By similar we mean SRI Indices of the same type (benchmark or tradable) with the same geographical scope.
[69] If we detect the presence of industry effects, we will add IP1, IP2 and IP3 to this regression in order to provide industry-adjusted results.

3.6.3 Regional Portfolios

We adopt a procedure similar to Lesser et al. (2014) by applying Fama-French (2015) model to a portfolio of SRI Indices. However, our method differs in some ways. First, we construct not only global but also regional portfolios. This allows analysing in more detail what kind of patterns are there in each region. Moreover, it permits seeing more clearly the differences that arise by analysing portfolios of SRI Indices instead of individual analysis. Second, our model is industry adjusted and therefore is free of possible industry-biases that could distort the results.

Each index has an equal weight in the corresponding portfolio. Since they come from three Series, this weight corresponds to 1/3.

Table 4 - Constitution of Regional Portfolios.
Source: own elaboration. We provide a list of the indices from each series used in the construction of each region portfolio.

Geographic Scope	FTSE4GOOD	MSCI ESG	STOXX ESG
Global	FTSE4GOOD GLB	MSCI WORLD ESG	STOXX GLOBAL ESG LEADERS
Europe	FTSE4GOOD EUR	MSCI EUROPE ESG	STOXX EU ESG LDRS 50
Japan	FTSE4GOOD Japan	MSCI Japan ESG	STOXX AP ESG LDRS 50
America	FTSE4GOOD US	MSCI NORTH AMERICA ESG	STOXX N.AMR ESG LDRS 50

3.6.4 Crisis Dummy

The fourth procedure consists of adding a dummy variable to all regressors of the Fama-French (2015) model to allow both alpha and betas to vary between crisis and non-crisis periods, resulting in the following model:

$$(3.9)\quad R_{it}\text{-}R_{ft} = \alpha_i + \beta_{1i}(R_{mt}\text{-}R_{ft}) + \beta_{2i}SML_t + \beta_{3i}HML_t + \beta_{4i}WML_t + \beta_{5i}RMW_t + \beta_{6i}CMA_t + \alpha_{2i}D_{C,t} + \beta_{7i}D_{Ct}(R_{mt}\text{-}R_f) + \beta_{8i}D_{Ct}SML_t + \beta_{9i}D_{Ct}HML_t + \beta_{10i}D_{Ct}WML_t + \beta_{11i}D_{Ct}RMW_t + \beta_{12i}D_{Ct}CMA_t + \varepsilon_{it}$$

Where $D_{C,t}$ represents a dummy variable that takes the value of 1 in periods of crisis and 0 in periods of non-crisis. Thus α_1 and β_{1-6} represent, respectively, the alpha and risk-exposures during non-crisis periods while α_2 and β_{7-12} represent, respectively, the difference of alpha and risk-exposures between crisis and non-crisis periods[70].

In the twentieth-first century there were essentially three of such moments: The dot com crisis of the early 2000's, the Global Financial Crisis and the European Sovereign Debt Crisis. We identify each of these periods as the longest time-span covered in the literature. Based on Table 19, which we already presented in the Review of the Literature, this results in the following periods:

Dot com crisis: January 2001 – December 2003

Global Financial Crisis: January 2007 – December 2009

Sovereign Debt Crisis: October 2009 – January 2013[71]

Since the earliest observations we include in our analysis only date back to 2004, our analysis only includes the Global and Sovereign Debt Crisis. We didn't choose any specific mentioned approach in Table 19 because, as can be easily deduced it could lead to even more different periods than the ones reported. Moreover, in our case the application of any of the

[70] If we detect the presence of industry effects, we will add IP1, IP2 and IP3 to this regression in order to provide industry-adjusted results.

[71] Since all SRI Indices are very correlated, we account for the Sovereign Debt Crisis on the analysis of all Indices and not just European Indices. Since the Global Financial Crisis and Sovereign Debt Crisis overlap, the two crisis are merged as a single period covering January 2007 and January 2013.

mentioned procedures would be far more complex, namely since we conduct individual analysis of a great set of SRI Indices. Thus, we adopt a simpler approach that avoids great discrepancies with the extant literature.

3.6.5 Variance Inflation Factor Analysis

As we have mentioned, our results are robust to problems of non-normality, heteroskedasticity and serial correlation. Nonetheless, we must test for the presence of additional common econometric problems, such as multicollinearity.

In order to test for potential multicollinearity problems, we followed the common approach used by authors such as Russo et al. (2016) and performed an analysis of the Variance Inflation Factors of each regressor in each individual regression. To that effect, we estimated OLS Regressions of the Fama-French (2015) model for each individual SRI Index. The sample period chosen for each regression was the same used in the separate GMM-System framework, which were, respectively, 02/12/2004 – 31/05/2017 for FTSE4GOOD Indices, 01/10/2007 – 31/05/2017 for MSCI ESG Indices and 28/07/2013 – 31/05/2017 for STOXX ESG Indices. We chose to only test Fama-French (2015) model since it contains all the regressors for the other models. Thus any conclusions about multicollinearity can be extended to CAPM, Fama-French (1993) and Carhart (1997) models[72].

The Variance Inflation Factor of a regressor measures the degree of inflation how much of the variance of a coefficient estimate is inflated due to the presence of multicollinearity. The centered variance of each regressor p is represented by the following formula:

$$\frac{1}{1-R_p^2} \tag{3.10}$$

Where R_p^2 represents the r squared of the regressor p on the remaining regressors.[73] Basing ourselves on the mentioned authors and others such as Gujaratti (2005), multicollinearity problems exist if we detect VIF values superior to 10.

[72] We don’t test the Fama-French (2015) model with industry factors. Since these factors result from a principal components analysis of the residuals of Fama-French (2015) regressions, they cannot be correlated with the respective regressors.

[73] Information retrieved from the support page of Eviews at the following website: http://www.eviews.com/help/helpintro.html#page/content/testing-Coefficient_Diagnostics.html.

3.7 Limitations

Despite our important contributions to the literature, some limitations must be acknowledged.

First, we were not able to include know SRI Indices in our analysis, namely Dow Jones Sustainability Indices and Calvert. However, it must be noted that these exclusions were motivated by the application of rigorous selection criteria of SRI Indices, which we explain and justify in appendix 2.

Second, it may be not possible to completely avoid inconsistencies in our analysis thanks to inherent differences between the three Series. There are differences in the geographic scope of some indices across the three Series. In particular, while both FTSE4GOOD Japan and MSCI ESG Japan exclusively cover Japan, STOXX ESG 50 Leaders covers other countries from theses region as well. MSCI ESG North America and STOXX ESG 50 North America both cover North America but FTSE4GOOD US is focused exclusively on the United States of America. However, the most important country in all cases continues to be respectively, Japan and the United States.

The types of indices also differ across the three Series. On the one hand, FTSE4GOOD Series have mostly tradable and benchmark indices for the same geographic scope, this is not the case with the MSCI ESG and STOXX ESG Series, which could create some difficulties in comparing SRI Indices. However, these are only restricted to the size factor since the only difference between tradable and SRI benchmark Indices is the fact the first choose companies the 50 or 100 largest companies of its conventional official benchmark.

4. Discussion of Results

4.1 Main Models

4.1.1 CAPM

We present the results from the application of the CAPM Model in table 5. As expected, almost all alphas are negative and not statistically significant, both by analysing individual p-values as well by looking at their respective J-statistics in each Series. Only two SRI Indices shows a statistically significant alpha at 5%, FTSE4GOOD Europe 50 and FTSE4GOOD Global 100 and both are also negative.

With respect to Beta, we can see differences between the three Series. While FTSE4GOOD Indices tend to be riskier than their conventional benchmarks, by showing a Beta superior to 1, the opposite is true for Indices from the MSCI ESG Series. There are different risk-patterns within the STOXX ESG Series. While indices from Asia Pacific and North America are less risky, European and Global Indices show more risk.

Moreover, within each Series, the highest Betas belong to Global SRI Indices. These results contrast with finance theory, which predicts that a higher degree of diversification would lead to less exposure to risk. However, it is coherent with recent empirical works found in literature as we have analysed in the Review of the Literature. Examples include authors such as Becchetti et al. (2015) and Leite et al. (2014) who found that SRI Funds with a more restricted geographical scope are not penalized in terms of diversification in relation to more global SRI Funds.

All SRI Indices reproduce very closely the movements of their official benchmarks, as can be seen by the high adjusted r-squared of each equation. We follow the approach from Schröder (2007) by conducting a spanning test for each regression to further examine whether an investor can expect an equal return and/or risk by investing in a SRI Index instead of its conventional benchmark. Thus, the null hypothesis for each regression is that alpha=0 and Beta=1. We found very similar results as the author by showing that, despite the high r squared of all indices, the null hypothesis is rejected at 10% level in most SRI Indices (11 out of 15). This provides evidence that SRI Indices, in spite of being very similar to their conventional counterparts, have a difference return and risk profile which further justified the application of more complex models.

Table 5 - CAPM Regressions for the FTSE4GOOD, MSCI ESG and STOXX ESG Indices

Source: Adapted from econometric software Eviews 9. We estimated the CAPM model in a GMM-System Framework for each of the three Series. Sample (observations): FTSE4GOOD - 02/12/2004 – 31/05/2017 (3260); MSCI ESG - 01/10/2007 – 31/05/2017 (2523); STOXX ESG - 28/07/2013 – 31/05/2017 (1048). Kernel: Bartlett, Bandwidth: Fixed (9), No prewhitening. Linear estimation after one-step weighting matrix. Results are rounded to three decimal places. We annualize alpha, as explained in Appendix 7. T-statistics are reported in parentheses. The variance-covariance matrix is robust for heteroskedasticity and autocorrelation as it was estimated using the Newey-West (1987) procedure. Since each System is exactly identified, the J-statistic of each system without additional restrictions is zero. J-stat (1) are J-statistics that serve to test the null hypothesis that the respective parameter is null throughout each system. We also conducted a spanning test for each regression, which test the null hypothesis that alpha=0 and Beta=1. The X2 statistic is reported in the last column. *** 1% significance ** 5% significance * 10% significance

SRI Index	**Alpha**	**Beta**	**Adj. R2**	**X2**
FTSE4GOOD EUR 50	-1.955%**	0.99***	0.980	9.152**
	(-2.359)	(182.658)		
FTSE4GOOD EUR	-0.422%	1.002***	0.993	1.661
	(-1.126)	(348.884)		
FTSE4GOOD GLB 100	-1.832%**	1.041***	0.948	42.152***
	(-2.149)	(159.758)		
FTSE4GOOD GLB	-0.702%	1.027***	0.962	20.363***
	(-1.244)	(171.403)		
FTSE4GOOD JAP	-0.204%	1.014***	0.985	18.997***
	(-0.296)	(318.379)		
FTSE4GOOD US 100	-0.278%	0.994***	0.968	0.919
	(-0.316)	(157.632)		
FTSE4GOOD US	-0.153%	0.997***	0.968	0.224
	(-0.184)	(158.403)		
J-stat (1)	7.604	139.381***		
MSCI EUROPE ESG	0.412%	0.981***	0.991	16.361***
	(0.662)	(205.555)		
MSCI JAPAN ESG	0.775%	0.857***	0.797	155.846***
	(0.276)	(74.518)		
MSCI NORTH AMERICA ESG	-0.269%	0.984***	0.992	12.446***
	(-0.491)	(214.6)		
MSCI WORLD ESG	-0.037%	0.992***	0.994	5.306*
	(-0.1)	(278.735)		
J-stat (1)	0.718	94.206***		
STOXX AP ESG LDRS 50	2.016%	0.92***	0.866	6.003**
	(0.834)	(27.72)		
STOXX EU ESG LDRS 50	-1.934%	1.065***	0.971	52.413***
	(-1.453)	(117.822)		
STOXX GLOBAL ESG LEADERS	-1.065%	1.108***	0.780	4.586
	(-0.418)	(21.514)		
STOXX N.AMR ESG LDRS 50	-0.553%	0.947***	0.900	17.070***
	(-0.282)	(73.154)		
J-stat (1)	2.956	71.089***		

4.1.2 Fama-French (1993)

Results from the application of the Fama-French (1993) model are summarized in table 6. In spite of the high R-squared of the CAPM equations, the addition of Fama-French (1993) factors proved to be relevant. Estimates of the SMB and HML factors are overall statistically significant both at the series and index level. No substantial changes were detected on alpha, even on FTSE4GOOD EUR 50 and FTSE4GOOD GLB 100 where it remains statistically significant and with a negative signal. However, now almost all Betas from the FTSE4GOOD Series are inferior to 1. This indicates that the higher risk-exposure of FTSE4GOOD indices is partly explained by the exposures to the size and value factors.

All Indices from the FTSE4GOOD Series show a statistically significant large cap bias. However, in the other series no discernible pattern can be detected, with differences in sign and statistical significance across different geographical scopes. As expected, tradable indices are less exposed to small caps compared with benchmark indices, as can be seen by the corresponding indices from the FTSE4GOOD Series and STOXX ESG Series. However, there are important differences. While in the FTSE4GOOD Series, tradable indices show a statistically significant large cap bias, in the STOXX ESG Series most tradable indices show a small cap bias, with the exception of STOXX EU ESG. These results could indicate that FTSE4GOOD may apply screening criteria more skewed towards larger companies while STOXX may be more tilted towards smaller companies.

By comparing SRI Indices with similar geographical scope, we can see different results by comparing the three Series. Starting by analysing Indices with a global scope, FTSE4Good GLB and FTSE4Good GLB 100 both show statistically significant large cap biases. MSCI World ESG shows a positive estimate but is not statistically significant. STOXX Global ESG Leaders shows a statistically significant small cap bias. Regional Indices also show no discernible pattern across the three Series. The only exception are indices which focus on Europe as they show a large cap bias (although it is not statistically significant in the case of MSCI Europe ESG).

With respect to the value factor HML, almost all SRI indices show a positive estimate (14 out of 15). These results could indicate that SRI Indices, in general, tend to contain more Value stocks than Growth stocks in relation to their conventional official benchmarks.

Table 6 - Fama-French (1993) Regressions for the FTSE4GOOD, MSCI ESG and STOXX ESG Indices
Source: Adapted from Econometric Software Eviews 9. We estimated the Fama-French (1993) model in a GMM-System Framework for each of the three Series. Sample (observations): FTSE4GOOD - 02/12/2004 – 31/05/2017 (3260); MSCI ESG - 01/10/2007 – 31/05/2017 (2523); STOXX ESG - 28/07/2013 – 31/05/2017 (1048). Kernel: Bartlett, Bandwidth: Fixed (9), No prewhitening. Linear estimation after one-step weighting matrix. Results are rounded to three decimal places. We annualize alpha, as explained in Appendix 7. T-statistics are reported in parentheses. The variance-covariance matrix is robust for heteroskedasticity and autocorrelation as it was estimated using the Newey-West (1987) procedure. Since each System is exactly identified, the J-statistic of each system without additional restrictions is zero. J-stat (1) are J-statistics that serve to test the null hypothesis that the respective parameter is null throughout each system. J-stat (2) represent the J-statistics that serve to test the null hypothesis that SMB and HML estimates are null throughout each system. *** 1% significance ** 5% significance * 10% significance.

SRI Index	**Alpha**	**Beta**	**SMB**	**HML**	**Adj. R2**
FTSE4GOOD EUR 50	-1.359%*	0.922***	-0.231***	0.024	0.984
	(-1.901)	(134.146)	(-17.618)	(1.307)	
FTSE4GOOD EUR	-0.246%	0.981***	-0.051***	0.054***	0.994
	(-0.726)	(325.956)	(-7.797)	(6.521)	
FTSE4GOOD GLB 100	-1.393%*	0.972***	-0.306***	0.067***	0.957
	(-1.898)	(133.064)	(-19.195)	(3.445)	
FTSE4GOOD GLB	-0.599%	1.002***	-0.078***	0.105***	0.964
	(-1.11)	(155.744)	(-4.863)	(5.632)	
FTSE4GOOD JAP	-0.135%	0.997***	-0.072***	0.061***	0.986
	(-0.215)	(347.046)	(-8.389)	(7.307)	
FTSE4GOOD US 100	-0.255%	0.998***	-0.084***	0.015	0.969
	(-0.297)	(138.102)	(-6.466)	(0.856)	
FTSE4GOOD US	-0.13%	0.999***	-0.06***	0.015	0.969
	(-0.158)	(142.905)	(-4.859)	(0.926)	
J-stat (1)	5.595	140.785***	118.146***	99.115***	
J-stat (2)	164.207***				
MSCI EUROPE ESG	0.48%	0.972***	-0.02	0.02	0.992
	(0.757)	(68.991)	(-1.062)	(0.503)	
MSCI JAPAN ESG	1.11%	0.844***	-0.062**	-0.043	0.797
	(0.399)	(64.752)	(-1.994)	(-1.382)	
MSCI NORTH AMERICA ESG	-0.241%	0.98***	0.028***	0.015*	0.992
	(-0.442)	(239.844)	(3.887)	(1.773)	
MSCI WORLD ESG	-0.012%	0.991***	0.008	0.03***	0.994
	(-0.033)	(246.344)	(1.07)	(2.939)	
J-stat (1)	0.906	100.189***	17.098***	9.883**	
J-stat (2)	34.783***				
STOXX AP ESG LDRS 50	1.935%	0.921***	0.028	0.07*	0.867
	(0.796)	(23.416)	(0.65)	(1.944)	
STOXX EU ESG LDRS 50	-0.569%	1.002***	-0.175***	0.115***	0.975
	(-0.479)	(91.934)	(-10.125)	(5.707)	
STOXX GLOBAL ESG LEADERS	-1.349%	1.16***	0.317***	0.299***	0.803
	(-0.563)	(22.144)	(5.32)	(6.403)	
STOXX N.AMR ESG LDRS 50	0.392%	0.945***	0.081***	0.208***	0.918
	(0.233)	(87.363)	(4.213)	(8.951)	
J-stat (1)	1.23	67.794***	58.585***	46.744***	
J-stat (2)	67.632***				

4.1.3 Carhart (1997)

Results from the application of the Carhart (1997) model are represented in table 7. At the series level, the inclusion of the momentum variable was statistically significant in all three Series, as can be seen by J-stat (1). At the index level, we can also see that this variable was statistically significant at 10% in most SRI Indices (11 out of 15). Moreover, its inclusion only caused minor modifications in some variables. In some indices, the estimate of HML factor changed sign, from positive to negative, but remained overall positive. However, HML estimates ceased to be statistically significant in most indices from the MSCI ESG Series, with the exception of MSCI ESG Global.

In most indices, despite the different patterns in terms of statistical significance, the estimates of the momentum variable tend to have a small magnitude and a negative sign (in 11 out of 15 indices). This implies that most SRI Indices are less invested in winner stocks and more in loser stocks in relation to their conventional Benchmarks. Nonetheless, there are different patterns within some Series, namely in indices belonging to MSCI ESG Series. Moreover, different patterns are also found when comparing indices with similar geographic scopes. The only exception are indices that cover the United States which share a contrarian bias.

Contrarian Bias appears to be less intense in tradable SRI indices. FTSE4GOOD EUR 50 and FTSE4GOOD GLB 100 show a positive but non-statistically significant momentum bias, in contrast to FTSE4GOOD EUR and FTSE4GOOD GLB which show negative coefficients. The magnitude of the contrarian bias of STOXX GLOBAL ESG is substantially higher than that of Regional indices from the STOXX ESG Series.

Table 7 - Carhart (1997) Regressions for the FTSE4GOOD, MSCI ESG and STOXX ESG Indices

Source: Adapted from Econometric Software Eviews 9. We estimated the Carhart (1997) model in a GMM-System Framework for each of the three Series. Sample (observations): FTSE4GOOD - 02/12/2004 – 31/05/2017 (3260); MSCI ESG - 01/10/2007 – 31/05/2017 (2523); STOXX ESG - 28/07/2013 – 31/05/2017 (1048). Kernel: Bartlett, Bandwidth: Fixed (9), No prewhitening. Linear estimation after one-step weighting matrix. Results are rounded to three decimal places. We annualize alpha, as explained in Appendix 7. T-statistics are reported in parentheses. The variance-covariance matrix is robust for heteroskedasticity and autocorrelation as it was estimated using the Newey-West (1987) procedure. Since each System is exactly identified, the J-statistic of each system without additional restrictions is zero. J-stat (1) are J-statistics that serve to test the null hypothesis that the respective parameter is null throughout each system. *** 1% significance ** 5% significance * 10% significance.

SRI Index	**Alpha**	**Beta**	**SMB**	**HML**	**WML**	**Adj. R2**
FTSE4GOOD EUR 50	-1.459%**	0.923***	-0.233***	0.029	0.012	0.984
	(-1.977)	(155.1)	(-16.622)	(1.395)	(0.933)	
FTSE4GOOD EUR	-0.174%	0.98***	-0.049***	0.05***	-0.008	0.994
	(-0.509)	(337.264)	(-7.059)	(5.043)	(-1.525)	
FTSE4GOOD GLB 100	-1.403%*	0.972***	-0.307***	0.069***	0.002	0.957
	(-1.879)	(138.49)	(-18.267)	(3.173)	(0.184)	
FTSE4GOOD GLB	-0.492%	0.999***	-0.072***	0.092***	-0.021**	0.964
	(-0.915)	(152.639)	(-4.446)	(4.738)	(-2.501)	
FTSE4GOOD JAP	-0.148%	0.997***	-0.065***	0.058***	-0.023***	0.987
	(-0.238)	(362.881)	(-7.565)	(7.054)	(-3.379)	
FTSE4GOOD US 100	-0.036%	0.99***	-0.08***	-0.032*	-0.07***	0.971
	(-0.043)	(133.825)	(-6.035)	(-1.881)	(-5.649)	
FTSE4GOOD US	0.086%	0.991***	-0.056***	-0.03*	-0.069***	0.970
	(0.108)	(137.563)	(-4.427)	(-1.916)	(-5.793)	
J-stat (1)	6.789	144.332***	112.344***	99.577***	38.534***	
MSCI EUROPE ESG	0.278%	0.978***	-0.024	0.033	0.033*	0.992
	(0.431)	(102.82)	(-1.26)	(0.777)	(1.673)	
MSCI JAPAN ESG	1.205%	0.846***	-0.067**	-0.041	0.017	0.797
	(0.431)	(65.593)	(-2.099)	(-1.299)	(0.757)	
MSCI NORTH AMERICA ESG	-0.237%	0.977***	0.028***	-0.001	-0.022***	0.992
	(-0.435)	(231.96)	(4.018)	(-0.092)	(-4.772)	
MSCI WORLD ESG	-0.001%	0.99***	0.009	0.026**	-0.006	0.994
	(-0.003)	(248.886)	(1.144)	(2.29)	(-1.055)	
J-stat (1)	0.576	111.474***	18.204***	11.424**	23.396***	
STOXX AP ESG LDRS 50	1.473%	0.93***	0.042	-0.007	-0.1***	0.870
	(0.621)	(21.978)	(0.91)	(-0.222)	(-2.993)	
STOXX EU ESG LDRS 50	-0.239%	1***	-0.174***	0.092***	-0.044***	0.976
	(-0.208)	(114.122)	(-10.829)	(4.724)	(-2.642)	
STOXX GLOBAL ESG LEADERS	-0.817%	1.148***	0.303***	0.169***	-0.174***	0.811
	(-0.352)	(25.931)	(5.39)	(3.358)	(-3.402)	
STOXX N.AMR ESG LDRS 50	0.444%	0.946***	0.066***	0.167***	-0.065***	0.921
	(0.269)	(95.196)	(3.523)	(8.145)	(-4.091)	
J-stat (1)	0.655	69.043***	57.908***	42.591***	19.776***	

4.1.4 Fama-French (2015)

Results from the application of the Fama-French (2015) model are summarised in table 8. By looking at the J-stat (1) of the profit and investment variables, it can be seen that the inclusion of these variables was statistically significant at 1% in all three Series. At the index level, however, results are very heterogeneous.

We can see that the inclusion of these variables did not substantially affect the estimates of the previous existing variables. The only notable change occurred with STOXX ESG Series, where most HML estimates became non-statistically significant. These results contrast with those of Fama-French (2015). In that paper, the authors found that the inclusion of profit and investment relatable variables (RMW and CMA) absorbed the effect of the Value Variable (HML). One possible explanation for such differences is the fact that the mentioned authors analysed conventional stocks while our focus is on SRI.

Within each Series, there are differences in terms of sign and statistical significance of both the RMW and CMA estimates. The same is true when comparing indices with similar geographical scope. However, it can be seen that most indices from the FTSE4GOOD Series show negative estimates of RMW, while indices from the MSCI ESG and STOXX ESG Series show the opposite pattern. In relation to estimates of the CMA Variable, most indices from all series show positive estimates. STOXX ESG Global stands out with a very high coefficient (0,42).

These results indicate that SRI Indices tend to select companies with a more conservative investment strategy. However, they have different profitability profiles according to which SRI agency was responsible for their construction. These results contrasts with Lesser et al. (2014) who also applied a similar model to Fama-French (2015) model. The authors applied it to a portfolio of Global SRI Indices and obtained non-statistically significant estimates for the profit and investment variables. Thus, these results the importance of studying SRI Indices individually and controlling for both the SRI provider and geographical scope to avoid ignoring important differences related to these factors, which end up cancelling each other and resulting in non-statistically significant results.

Table 8 - Fama-French (2015) Regressions for the FTSE4GOOD, MSCI ESG and STOXX ESG Indices
Source: Adapted from Econometric Software Eviews 9. We estimated the Fama-French (2015) model in a GMM-System Framework for each of the three Series. Sample (observations): FTSE4GOOD - 02/12/2004 – 31/05/2017 (3260); MSCI ESG - 01/10/2007 – 31/05/2017 (2523); STOXX ESG - 28/07/2013 – 31/05/2017 (1048). The variance-covariance matrix is robust for heteroskedasticity and autocorrelation as it was estimated using the Newey-West (1987) procedure. Kernel: Bartlett, Bandwidth: Fixed (9), No prewhitening. Linear estimation after one-step weighting matrix. Results are rounded to three decimal places. We annualize alpha, as explained in Appendix 4. T-statistics are reported in parentheses. J-stat (1) are J-statistics that serve to test the null hypothesis that the respective parameter is null throughout each system. J-stat (2) represent the J-statistics that serve to test the null hypothesis that RMW and CMA estimates are null throughout each system. *** 1% significance ** 5% significance * 10% significance.

SRI Index	**Alpha**	**Beta**	**SMB**	**HML**	**WML**
FTSE4GOOD EUR 50	-1.756%**	0.924***	-0.227***	0.051**	0.006
	(-2.322)	(165.4)	(-18.459)	(2.096)	(0.583)
FTSE4GOOD EUR	-0.044%	0.978***	-0.053***	0.047***	-0.005
	(-0.125)	(276.035)	(-7.568)	(2.842)	(-0.864)
FTSE4GOOD GLB 100	-1.379%*	0.981***	-0.303***	0.026	-0.002
	(-1.83)	(110.276)	(-17.564)	(1.017)	(-0.178)
FTSE4GOOD GLB	-0.65%	1.01***	-0.065***	0.067***	-0.029***
	(-1.195)	(120.987)	(-3.989)	(2.985)	(-3.213)
FTSE4GOOD JAP	-0.082%	0.995***	-0.065***	0.058***	-0.018***
	(-0.132)	(363.036)	(-7.591)	(6.011)	(-2.664)
FTSE4GOOD US 100	0.579%	0.99***	-0.113***	-0.122***	-0.054***
	(0.734)	(145.175)	(-10.507)	(-9.018)	(-5.557)
FTSE4GOOD US	0.651%	0.991***	-0.086***	-0.116***	-0.054***
	(0.861)	(144.427)	(-8.354)	(-8.889)	(-5.771)
J-stat (1)	11.956	171.166***	121.804***	106.704***	41.515***
J-stat (2)	123.881***				
MSCI EUROPE ESG	-0.065%	0.973***	-0.02	0.079	0.03**
	(-0.091)	(107.238)	(-1.522)	(1.407)	(2.472)
MSCI JAPAN ESG	1.195%	0.846***	-0.066**	-0.047	0.015
	(0.425)	(65.211)	(-2.084)	(-1.366)	(0.579)
MSCI NORTH AMERICA ESG	-0.344%	0.981***	0.029***	-0.013	-0.024***
	(-0.64)	(198.851)	(3.749)	(-1.369)	(-5.136)
MSCI WORLD ESG	-0.372%	0.999***	0.019***	0.031**	-0.014**
	(-1.047)	(250.212)	(2.738)	(2.314)	(-2.547)
J-stat (1)	1.82	133.211***	21.517***	21.449***	24.809***
J-stat (2)	59.749***				
STOXX AP ESG LDRS 50	1.031%	0.936***	0.04	0.052	-0.108***
	(0.431)	(19.728)	(0.824)	(1.537)	(-2.791)
STOXX EU ESG LDRS 50	0.702%	0.996***	-0.188***	0.012	-0.041***
	(0.642)	(147.436)	(-12.19)	(0.432)	(-2.866)
STOXX GLOBAL ESG LEADERS	-1.529%	1.201***	0.351***	0.037	-0.204***
	(-0.647)	(27.04)	(6.179)	(0.546)	(-3.853)
STOXX N.AMR ESG LDRS 50	0.492%	0.965***	0.091***	0.111***	-0.065***
	(0.309)	(94.73)	(5.269)	(3.679)	(-3.971)
J-stat (1)	1.741	82.401***	62.99***	14.494***	22.74***
J-stat (2)	51.65***				

SRI Index	RMW	CMA	Adj. R2
FTSE4GOOD EUR 50	0.071***	0.009	0.984
	(2.588)	(0.388)	
FTSE4GOOD EUR	-0.025	-0.022*	0.994
	(-1.401)	(-1.813)	
FTSE4GOOD GLB 100	-0.048*	0.091***	0.957
	(-1.81)	(2.742)	
FTSE4GOOD GLB	0.009	0.085***	0.964
	(0.34)	(2.842)	
FTSE4GOOD JAP	-0.03**	-0.023	0.987
	(-2.178)	(-1.459)	
FTSE4GOOD US 100	-0.233***	0.13***	0.975
	(-10.977)	(5.947)	
FTSE4GOOD US	-0.217***	0.127***	0.974
	(-10.232)	(5.785)	
J-stat (1)	83.14***	56.095***	
J-stat (2)			
MSCI EUROPE ESG	0.099	-0.05**	0.992
	(1.522)	(-2.285)	
MSCI JAPAN ESG	-0.01	0.011	0.797
	(-0.181)	(0.234)	
MSCI NORTH AMERICA ESG	0.005	0.042***	0.992
	(0.378)	(2.865)	
MSCI WORLD ESG	0.075***	0.045***	0.994
	(6.517)	(2.777)	
J-stat (1)	43.788***	17.153***	
J-stat (2)			
STOXX AP ESG LDRS 50	0.177***	0.04	0.871
	(3.752)	(0.598)	
STOXX EU ESG LDRS 50	-0.2***	-0.073**	0.977
	(-5.364)	(-2.236)	
STOXX GLOBAL ESG LEADERS	0.163	0.418***	0.817
	(1.56)	(4.109)	
STOXX N.AMR ESG LDRS 50	0.084**	0.168***	0.924
	(2.481)	(4.089)	
J-stat (1)	27.513***	29.601***	
J-stat (2)			

4.2 Robustness Tests

4.2.1 Industry Adjusted Regressions

Results from the application of the Fama-French (2015) model with industry factors are reported in table 9. The inclusion of industry factors allows checking whether pre-existing bias are justified or not by industry exposures and provides more robust results if such biases are detected. As authors such as Derwall et al. (2005) and Erragragui et al. (2016) have pointed out, it's difficult to interpret the coefficients of principal components. However, at the series level, their inclusion was statistically significant both at the series and index level. These results confirm the existence of industry effects and, therefore, justify the importance of conducting industry-adjusted analyses on the returns of SRI Indices.

Most alphas remained non-statistically significant but now most exhibit positive signal. However, there were different impacts at the individual level, especially in the FTSE4GOOD Series, where they are statistically significant at the series level. American indices now show a positive and statistically significant alpha. FTSE4GOOD EUR 50 remained with a statistically significant negative alpha while in the case of FTSE4GOOD GLB 100 it ceased to be statistically significant. In relation to Beta, now all the Global SRI Indices have an estimate superior to 1 while regional SRI Indices show a beta inferior to 1. These results exacerbate those already found and explained in the application of the CAPM model. A few changes were detected in the magnitude of the SMB estimates. Large cap biases in indices from the FTSE4GOOD Series became more prominent. However, in the other series no substantial changes were detected.

With respect to the value factor, some important changes were detected. HML estimates of global indices either reversed sign, such as in the case of FTSE4GOOD and STOXX ESG, or became non-statistically significant, such as in the case of MSCI WORLD ESG. No significant changes were detected with respect to the momentum factor. Almost all indices continue to show a negative momentum bias. With respect to estimates of RMW, there were changes from positive to negative across the three series. In particular, there was a change of signal in some regressions pertaining to an American index (MSCI ESG Series) and global indices (FTSE4GOOD and STOXX ESG Series). In the latter, the most important change occurred with STOXX GLOBAL ESG which showed a drastic change from a non-statistically positive coefficient to a statistically significant negative coefficient. Finally, in the case of CMA, more indices now show negative estimates, resulting in increasing different patterns within each Series and Indices with a similar geographic scope. The only exception are European and global indices, which show, respectively, a negative and positive estimate across the three series.

Table 9 - Industry-Adjusted Fama-French (2015) Regressions for the FTSE4GOOD, MSCI ESG and STOXX ESG Indices

Source: Adapted from Econometric Software Eviews 9. We estimated the Fama-French (2015) model in a GMM-System Framework for each of the three Series and adjusted for industry effects by adding industry factors. Sample (observations): FTSE4GOOD - 02/12/2004 – 31/05/2017 (3145); MSCI ESG - 01/10/2007 – 31/05/2017 (2434); STOXX ESG - 28/07/2013 – 31/05/2017 (1011). The variance-covariance matrix is robust for heteroskedasticity and autocorrelation as it was estimated using the Newey-West (1987) procedure. Kernel: Bartlett, Bandwidth: Fixed (9), No prewhitening. Linear estimation after one-step weighting matrix. We annualize alpha, as explained in Appendix 7. T-statistics are reported in parentheses. Since each System is exactly identified, the J-statistic of each system without additional restrictions is zero. J-stat (1) are J-statistics that serve to test the null hypothesis that the respective parameter is null throughout each system. J-stat (2) represent the J-statistics that serve to test the null hypothesis that IP1, IP2 and IP3 estimates are null throughout each system. *** 1% significance ** 5% significance * 10% significance.

SRI Index	**Alpha**	**Beta**	**SMB**	**HML**	**WML**
FTSE4GOOD EUR 50	-1.728%**	0.905***	-0.254***	0.08***	0.004
	(-2.319)	(163.767)	(-19.588)	(3.711)	(0.404)
FTSE4GOOD EUR	0.007%	0.973***	-0.063***	0.05***	-0.005
	(0.02)	(229.877)	(-8.613)	(3.278)	(-0.909)
FTSE4GOOD GLB 100	-0.251%	0.979***	-0.342***	-0.044**	0.008
	(-0.359)	(123.858)	(-19.656)	(-1.991)	(0.735)
FTSE4GOOD GLB	0.487%	1.009***	-0.107***	-0.019	-0.021***
	(1.063)	(148.264)	(-7.239)	(-0.972)	(-2.943)
FTSE4GOOD JAP	0.043%	0.992***	-0.067***	0.056***	-0.017**
	(0.064)	(332.873)	(-8.596)	(5.913)	(-2.497)
FTSE4GOOD US 100	1.345%**	0.994***	-0.139***	-0.107***	-0.028***
	(2.048)	(188.119)	(-15.735)	(-10.356)	(-4.789)
FTSE4GOOD US	1.464%**	0.996***	-0.114***	-0.104***	-0.028***
	(2.369)	(180.716)	(-12.426)	(-10.645)	(-5.106)
J-stat (1)	16.425**	161.412***	118.848***	98.006***	49.137***
J-stat (2)	166.133***				
MSCI EUROPE ESG	-0.168%	0.98***	-0.016	0.067	0.031***
	(-0.237)	(100.629)	(-1.114)	(1.187)	(2.785)
MSCI JAPAN ESG	1.931%	0.809***	-0.047	-0.033	0.024
	(0.689)	(54.444)	(-1.46)	(-0.96)	(0.916)
MSCI NORTH AMERICA ESG	-0.237%	0.988***	0.03***	-0.019**	-0.021***
	(-0.418)	(268.436)	(3.823)	(-2.063)	(-5.024)
MSCI WORLD ESG	-0.052%	1***	0.007	0.002	-0.012**
	(-0.141)	(289.268)	(0.896)	(0.149)	(-2.328)
J-stat (1)	1.153	124.051***	13.756***	9.11*	26.988***
J-stat (2)	89.153***				
STOXX AP ESG LDRS 50	2.04%	0.931***	0.028	0.04	-0.111***
	(0.827)	(17.185)	(0.563)	(1.16)	(-2.659)
STOXX EU ESG LDRS 50	0.723%	0.99***	-0.195***	0.033	-0.05***
	(0.656)	(147.632)	(-12.361)	(1.249)	(-3.46)
STOXX GLOBAL ESG LEADERS	-0.123%	1.167***	0.185***	-0.225***	-0.182***
	(-0.061)	(34.176)	(3.786)	(-3.785)	(-4.087)
STOXX N.AMR ESG LDRS 50	-0.089%	1***	0.146***	0.097***	-0.034**
	(-0.057)	(76.748)	(7.85)	(3.097)	(-2.111)
J-stat (1)	1.436	84.407***	65.17***	21.431***	21.267***
J-stat (2)	75.825***				

SRI Index	RMW	CMA	IP1	IP2	IP3	Adj. R2
FTSE4GOOD EUR 50	0.097***	0.002	-0.011***	-0.005**	0.006*	0.985
	(4.055)	(0.078)	(-6.521)	(-2.037)	(1.811)	
FTSE4GOOD EUR	-0.02	-0.03**	-0.003***	-0.006***	-0.004**	0.994
	(-1.143)	(-2.359)	(-3.612)	(-4.51)	(-2.093)	
FTSE4GOOD GLB 100	-0.152***	0.028	-0.011***	-0.007**	-0.04***	0.961
	(-5.768)	(0.829)	(-4.744)	(-2.108)	(-6.55)	
FTSE4GOOD GLB	-0.132***	0.031	-0.019***	-0.004	-0.038***	0.971
	(-6.005)	(1.066)	(-8.563)	(-1.23)	(-8.497)	
FTSE4GOOD JAP	-0.03**	-0.023	-0.002*	-0.003*	0.002	0.987
	(-2.13)	(-1.457)	(-1.95)	(-1.876)	(0.768)	
FTSE4GOOD US 100	-0.219***	-0.012	0.006***	-0.038***	-0.063***	0.983
	(-16.143)	(-0.772)	(5.267)	(-16.343)	(-17.957)	
FTSE4GOOD US	-0.208***	-0.012	0.006***	-0.035***	-0.065***	0.982
	(-15.896)	(-0.777)	(5.371)	(-16.029)	(-18.018)	
J-stat (1)	98.541***	27.632***	48.525***	107.233***	114.607***	
J-stat (2)						
MSCI EUROPE ESG	0.099	-0.054***	0.003**	-0.013***	-0.01***	0.993
	(1.482)	(-2.667)	(2.546)	(-5.363)	(-3.492)	
MSCI JAPAN ESG	-0.023	-0.028	-0.04***	0.021**	-0.011	0.807
	(-0.445)	(-0.597)	(-7.058)	(2.179)	(-0.849)	
MSCI NORTH AMERICA ESG	-0.001	0.04***	-0.003*	-0.001	-0.011***	0.992
	(-0.057)	(2.632)	(-1.888)	(-0.414)	(-4.033)	
MSCI WORLD ESG	0.038***	0.028*	-0.005***	-0.004***	-0.013***	0.995
	(2.75)	(1.777)	(-5.813)	(-3.274)	(-6.442)	
J-stat (1)	13.328***	15.283***	63.484***	28.618***	27.017***	
J-stat (2)						
STOXX AP ESG LDRS 50	0.164***	0.017	-0.006	0.023***	-0.001	0.871
	(3.473)	(0.239)	(-0.915)	(3.308)	(-0.12)	
STOXX EU ESG LDRS 50	-0.188***	-0.078**	-0.003	-0.014***	-0.009**	0.977
	(-5.221)	(-2.433)	(-1.206)	(-3.908)	(-1.987)	
STOXX GLOBAL ESG LEADERS	-0.304***	0.307***	-0.099***	0.041***	-0.053***	0.893
	(-3.06)	(4.239)	(-16.613)	(4.648)	(-4.437)	
STOXX N.AMR ESG LDRS 50	0.164***	0.175***	-0.028***	0.023***	0.019***	0.929
	(4.072)	(4.22)	(-6.161)	(3.31)	(2.808)	
J-stat (1)	35.041***	28.337***	67.638***	40.864***	33.529***	
J-stat (2)						

4.2.2 Single GMM System

We present the results of estimation of the Fama-French (2015) model with industry factors in a single GMM System framework in table 10. This approach allows to make more robust comparisons between indices from different series. This is accomplished by taking into account industry effects, which were already proven to exist in three GMM system framework, and by analysing a common sample period (28/07/2013 - 31/05/2017).

Now most alphas show positive estimates, despite continuing to be overall non-statistically significant, which that performance results are sensitive to the period of analysis. FTSE4GOOD Indices continue to show a statistically significant large cap bias while most indices from MSCI ESG and STOXX ESG Series show a small cap bias, although not all are statistically significant. With respect to HML, now most indices from FTSE4GOOD and MSCI ESG Series show a statistically significant growth bias. These results contrast with the different patterns already found and analysed in the STOXX ESG Series.

The momentum factor estimates experienced some minor changes in some indices, namely in signal. One notable change occurred with MSCI ESG Japan which now shows a statistically significant momentum bias which is also the largest of all indices. However, almost all indices continue to show negative estimates. Similar results are found with respect to RMW and CMA estimates.

Overall, this analysis confirms most of the previous results. There are different patterns in many biases within indices belonging to the same Series. This suggests that there are important differences in characteristics of selected companies between different countries, even when applying similar screening procedures. However, the fact that there are no clear patterns when comparing indices with a similar geographical scope from different series suggests that there are differences in screening criteria.

Table 10 - Single GMM System estimation of Industry-Adjusted Fama-French (2015) Regressions for the FTSE4GOOD, MSCI ESG and STOXX ESG Indices

Source: Adapted from econometric software Eviews 9. We estimated the Fama-French (2015) model in a Single GMM-System Framework of the series. Sample (observations): 28/07/2013 – 31/05/2017 (1011). The variance-covariance matrix is robust for heteroskedasticity and autocorrelation as it was estimated using the Newey-West (1987) procedure. Kernel: Bartlett, Bandwidth: Fixed (9), No prewhitening. Linear estimation after one-step weighting matrix. Results are rounded to three decimal places. We annualize alpha, as explained in Appendix 7. T-statistics are reported in parentheses. Since each System is exactly identified, the J-statistic of each system without additional restrictions is zero. J-stat (1) are J-statistics that serve to test the null hypothesis that the respective parameter is null throughout each system. J-stat (2) represent the J-statistics that serve to test the null hypothesis that IP1, IP2 and IP3 estimates are null throughout each system. *** 1% significance ** 5% significance * 10% significance.

SRI Index	**Alpha**	**Beta**	**SMB**	**HML**	**WML**
FTSE4GOOD EUR 50	0.038%	0.91***	-0.24***	-0.094***	-0.06***
	(0.036)	(66.194)	(-10.342)	(-2.733)	(-5.015)
FTSE4GOOD EUR	0.457%	0.993***	-0.052***	-0.063***	-0.005
	(1.159)	(367.07)	(-7.893)	(-4.361)	(-1.28)
FTSE4GOOD GLB 100	0.761%	0.977***	-0.279***	-0.138***	-0.031**
	(0.762)	(116.036)	(-18.599)	(-4.901)	(-2.275)
FTSE4GOOD GLB	0.344%	1.026***	-0.044***	-0.112***	-0.041***
	(0.589)	(135.819)	(-3.973)	(-6.02)	(-4.303)
FTSE4GOOD JAP	0.284%	0.997***	-0.056***	0.067***	0.012
	(0.32)	(183.803)	(-4.968)	(4.87)	(1.256)
FTSE4GOOD US 100	1.477%	0.988***	-0.081***	-0.022	-0.004
	(1.504)	(145.955)	(-6.399)	(-1.167)	(-0.415)
FTSE4GOOD US	1.583%*	0.993***	-0.045***	-0.026	-0.01
	(1.806)	(172.152)	(-4.073)	(-1.637)	(-1.202)
MSCI EUROPE ESG	-0.196%	0.999***	0.028**	-0.121***	0.027***
	(-0.291)	(137.398)	(2.241)	(-6.379)	(3.579)
MSCI JAPAN ESG	1.133%	0.747***	-0.02	-0.008	0.106**
	(0.3)	(24.03)	(-0.344)	(-0.118)	(2.401)
MSCI NORTH AMERICA ESG	-0.907%*	1.002***	0.029***	-0.041***	-0.016***
	(-1.662)	(238.337)	(4.26)	(-3.865)	(-2.764)
MSCI WORLD ESG	-0.828%**	1.005***	0.044***	-0.047***	-0.016***
	(-2.133)	(269.043)	(6.687)	(-4.451)	(-2.984)
STOXX AP ESG LDRS 50	2.04%	0.931***	0.028	0.04	-0.111***
	(0.827)	(17.185)	(0.563)	(1.16)	(-2.659)
STOXX EU ESG LDRS 50	0.723%	0.99***	-0.195***	0.033	-0.05***
	(0.656)	(147.632)	(-12.361)	(1.249)	(-3.46)
STOXX GLOBAL ESG LEADERS	-0.123%	1.167***	0.185***	-0.225***	-0.182***
	(-0.061)	(34.176)	(3.786)	(-3.785)	(-4.087)
STOXX N.AMR ESG LDRS 50	-0.089%	1***	0.146***	0.097***	-0.034**
	(-0.057)	(76.748)	(7.85)	(3.097)	(-2.111)
J-stat (1)	17.165	100.082***	95.818***	68.48***	73.184***
J-stat (2)	115.697***				

SRI Index	RMW	CMA	IP1	IP2	IP3	Adj. R2
FTSE4GOOD EUR 50	-0.018	0.135***	-0.012***	0.001	0.012***	0.981
	(-0.41)	(4.437)	(-4.482)	(0.145)	(2.896)	
FTSE4GOOD EUR	-0.097***	0.026**	-0.001	-0.002	0.003	0.996
	(-5.965)	(1.996)	(-1.439)	(-0.982)	(1.196)	
FTSE4GOOD GLB 100	-0.162***	-0.04	-0.02***	0.002	-0.047***	0.971
	(-4.691)	(-1.106)	(-8.213)	(0.5)	(-9.942)	
FTSE4GOOD GLB	-0.158***	0.062***	-0.032***	-0.005*	-0.041***	0.988
	(-6.247)	(2.672)	(-18.378)	(-1.745)	(-13.2)	
FTSE4GOOD JAP	-0.026	-0.018	0.004*	-0.003	-0.003	0.990
	(-1.049)	(-0.879)	(1.831)	(-1.085)	(-0.78)	
FTSE4GOOD US 100	-0.058**	-0.085***	0.002	-0.025***	-0.055***	0.974
	(-2.442)	(-2.879)	(0.675)	(-5.74)	(-11.536)	
FTSE4GOOD US	-0.054***	-0.05*	0.003	-0.026***	-0.055***	0.979
	(-2.626)	(-1.953)	(1.347)	(-7.172)	(-13.167)	
MSCI EUROPE ESG	-0.026	-0.063***	0.002	-0.011***	-0.005*	0.992
	(-1.072)	(-3.384)	(1.116)	(-5.143)	(-1.752)	
MSCI JAPAN ESG	-0.095	-0.055	-0.084***	0.025*	0.016	0.793
	(-1.037)	(-0.525)	(-7.074)	(1.827)	(0.776)	
MSCI NORTH AMERICA ESG	0.02	0.133***	0	-0.004	0.004	0.991
	(1.347)	(9.369)	(-0.356)	(-1.488)	(1.262)	
MSCI WORLD ESG	0.055***	0.102***	0.001	-0.006***	0.004**	0.994
	(4.811)	(6.867)	(1.227)	(-3.384)	(2.041)	
STOXX AP ESG LDRS 50	0.164***	0.017	-0.006	0.023***	-0.001	0.871
	(3.473)	(0.239)	(-0.915)	(3.308)	(-0.12)	
STOXX EU ESG LDRS 50	-0.188***	-0.078**	-0.003	-0.014***	-0.009**	0.977
	(-5.221)	(-2.433)	(-1.206)	(-3.908)	(-1.987)	
STOXX GLOBAL ESG LEADERS	-0.304***	0.307***	-0.099***	0.041***	-0.053***	0.893
	(-3.06)	(4.239)	(-16.613)	(4.648)	(-4.437)	
STOXX N.AMR ESG LDRS 50	0.164***	0.175***	-0.028***	0.023***	0.019***	0.929
	(4.072)	(4.22)	(-6.161)	(3.31)	(2.808)	
J-stat (1)	59.696***	63.854***	98.844***	76.534***	83.003***	
J-stat (2)						

4.2.2 Regional Portfolios

We follow a similar procedure to Lesser et al. (2014) and apply a 5 factor Fama-French model to portfolios of SRI Indices with a similar geographic scope. Results are represented in table 11. However, our procedure has two main differences. First, it adjusts for the industry effects. Second, we do construct portfolios not only with a global scope but also with a regional coverage. This allows to analyse in more detail differences between countries. Each portfolio pertaining to a specific geographic scope is constructed with an equal-weighted position in a corresponding index from each of the three series.

It's necessary to interpret with caution the estimates of each portfolio. As we have mentioned, the simplification of the analysis comes at the expense of agglutinating the characteristics of individual indices which, as we have seen, differ extensively even between those with similar geographic scopes. This can be seen by the fact that none of the alphas of the portfolios are statistically significant while in the previous analysis it was possible to find indices with a statistically alpha in the FTSE4GOOD and MSCI ESG Series. However, this procedure allows to confirm some of our previous findings. Overall, alphas continue not to be statistically significant but show different signals. Global portfolios show a beta higher than 1 while regional portfolios show the opposite results. Risk-exposures are statistically significant in all four portfolios but differ considerably across different geographic scopes. The only exception is the contrarian bias found in all portfolios, which was also found previously in most individual indices.

However, some findings contrast with the literature, such as the small cap bias associated with the United States and the large cap with Europe. Since this pattern did not clearly emerge in previous analysis, these results reinforce our argument that the construction of portfolios may agglutinate the characteristics of individual indices and must be interpreted with caution.

Table 11 - Industry-Adjusted Single GMM System estimation of Fama-French (2015) Regressions Index Portfolios

Source: Adapted from econometric software Eviews 9. We estimated the Fama-French (2015) model adjusted for industry effects in a Single GMM-System Framework for Portfolios composed of indices with a similar geographical scope. Sample (observations): 28/07/2013 – 31/05/2017 (1011). Kernel: Bartlett, Bandwidth: Fixed (9), No prewhitening. Linear estimation after one-step weighting matrix. Results are rounded to three decimal places. We annualize alpha, as explained in Appendix 7. T-statistics are reported in parentheses. The variance-covariance matrix is robust for heteroskedasticity and autocorrelation as it was estimated using the Newey-West (1987) procedure. Since each System is exactly identified, the J-statistic of each system without additional restrictions is zero. J-stat (1) are J-statistics that serve to test the null hypothesis that the respective parameter is null throughout each system. *** 1% significance ** 5% significance * 10% significance.

Portfolio	Alpha	Beta	SMB	HML	WML	RMW	CMA
Global	-0.241%	1.071***	0.051***	-0.14***	-0.078***	-0.139***	0.177***
	(-0.293)	(73.945)	(2.615)	(-5.643)	(-4.29)	(-3.31)	(5.883)
Europe	0.436%	0.997***	-0.081***	-0.018*	-0.017***	-0.104***	-0.021*
	(1.039)	(367.411)	(-13.188)	(-1.791)	(-3.085)	(-6.979)	(-1.651)
Japan	1.025%	0.884***	-0.058***	0.045**	-0.042***	0.117***	-0.036
	(0.747)	(115.825)	(-4.122)	(2.276)	(-3.293)	(4.574)	(-1.371)
America	-0.137%	0.997***	0.061***	0.027**	-0.021***	0.06***	0.076***
	(-0.198)	(205.512)	(7.27)	(2.258)	(-2.988)	(3.448)	(4.739)
J-stat (1)	2.286	80.02***	69.248***	32.5***	21.483***	38.541***	37.648***

Portfolio	IP1	IP2	IP3	Adj. R2
Global	-0.043***	0.01***	-0.029***	0.975
	(-17.712)	(2.715)	(-6.039)	
Europe	-0.001	-0.006***	-0.003	0.997
	(-0.85)	(-3.952)	(-1.506)	
Japan	-0.011***	0.012***	0	0.976
	(-4.332)	(3.078)	(0.004)	
America	-0.015***	0.001	-0.008**	0.986
	(-9.971)	(0.474)	(-2.512)	
J-stat (1)	73.549***	26.065***	29.307***	

4.2.3 Crisis Dummy

Results of the application of the Fama-French (2015) model with a crisis dummy are represented in table 12. Overall most alphas are not statistically significant and mostly negative. However, there was a shift in the FTSE4GOOD Series. Alphas of FTSE4GOOD EUR 50 and FTSE4GOOD GLB 100 continue to be negative but no longer statistically significant while the opposite effect is seen with FTSE4GOOD US. There's no evidence of time-varying alpha but most alphas are positive, which is coherent with results found in the literature which point out that SRI tends to outperform Conventional Investment in periods of market downturns.

In terms of Beta, most statistically significant estimates of DC.Beta are negative. However there are different patterns within each Series and in indices with similar geographic scope. The only exceptions are Japanese indices which show a statistically significant positive estimate in both series.

Both Series show a growth bias during non-crisis periods which shifts into a value bias during crisis which is consistent with results obtained by other authors such as Scholtens (2005). Most WML estimates are negative during non-crisis periods but tend to go in the opposite direction during crisis and are not statistically significant at the series level in the case of MSCI ESG. Finally, both RMW and CMA estimates show different patterns in both periods.

Table 12 - Fama-French (2015) Regressions with a Crisis Dummy for the FTSE4GOOD, MSCI ESG and STOXX ESG Indices

Source: Adapted from econometric software Eviews 9. We estimated the Fama-French (2015) model with in a GMM-System Framework for FTSE4GOOD and MSCI ESG Series with a crisis dummy that comprises January 2007 - January 2013. STOXX ESG was not analysed because its period of observations did not contain any crisis periods. We also adjust our analysis for the presence of industry effects. Sample (observations): FTSE4GOOD - 02/12/2004 – 31/05/2017 (3260); MSCI ESG - 01/10/2007 – 31/05/2017 (2523). The variance-covariance matrix is robust for heteroskedasticity and autocorrelation as it was estimated using the Newey-West (1987) procedure. Kernel: Bartlett, Bandwidth: Fixed (9), No prewhitening. Linear estimation after one-step weighting matrix. Results are rounded to three decimal places. We annualize alpha, as explained in Appendix 7. T-statistics are reported in parentheses. Since each System is exactly identified, the J-statistic of each system without additional restrictions is zero. J-stat (1) are J-statistics that serve to test the null hypothesis that the respective parameter is null throughout each system. J-stat (2) represent the J-statistics that serve to test the null hypothesis that the estimates of the dummy variables are null throughout each system. *** 1% significance ** 5% significance * 10% significance.

SRI Index	Alpha	Beta	SMB	HML	WML
FTSE4GOOD EUR 50	-1.063%	0.926***	-0.217***	-0.046*	-0.044***
	(-1.329)	(83.552)	(-11.656)	(-1.757)	(-3.446)
FTSE4GOOD EUR	-0.07%	0.99***	-0.051***	-0.045***	-0.006*
	(-0.216)	(399.962)	(-8.947)	(-4.174)	(-1.795)
FTSE4GOOD GLB 100	0.145%	0.979***	-0.262***	-0.11***	-0.05***
	(0.183)	(128.968)	(-18.313)	(-4.503)	(-4.205)
FTSE4GOOD GLB	0.315%	1.02***	-0.043***	-0.072***	-0.039***
	(0.616)	(132.49)	(-3.783)	(-3.69)	(-3.768)
FTSE4GOOD JAP	-0.238%	0.984***	-0.057***	0.099***	0.005
	(-0.295)	(235.454)	(-6.391)	(7.401)	(0.391)
FTSE4GOOD US 100	1.219%	0.986***	-0.109***	-0.061***	-0.044***
	(1.622)	(178.857)	(-11.851)	(-5.262)	(-5.877)
FTSE4GOOD US	1.472%**	0.992***	-0.068***	-0.061***	-0.042***
	(2.175)	(210.719)	(-8.247)	(-5.866)	(-6.549)
J-stat (1)	13.712*	117.185***	112.649***	62.853***	60.474***
J-stat (2)	167.488***				
MSCI EUROPE ESG	0.078%	0.999***	0.025**	-0.124***	0.03***
	(0.122)	(146.535)	(1.989)	(-6.866)	(4.067)
MSCI JAPAN ESG	-0.326%	0.774***	-0.026	-0.013	0.062
	(-0.085)	(28.871)	(-0.521)	(-0.215)	(1.347)
MSCI NORTH AMERICA ESG	-0.709%	1.01***	0.028***	-0.058***	-0.015***
	(-1.325)	(257.21)	(4.086)	(-5.938)	(-2.622)
MSCI WORLD ESG	-0.35%	1.005***	0.028***	-0.085***	-0.015***
	(-0.861)	(286.428)	(4.293)	(-7.737)	(-2.853)
J-stat (1)	1.801	69.747***	22.615***	50.494***	27.527***
J-stat (2)	99.597***				

SRI Index	**RMW**	**CMA**	**DC**	**DC.Beta**	**DC.SMB**
FTSE4GOOD EUR 50	0.025	0.085***	-0.003	-0.033***	-0.064***
	(0.657)	(3.447)	(-0.542)	(-2.77)	(-2.748)
FTSE4GOOD EUR	-0.083***	0.011	0.001	-0.03***	-0.024**
	(-6.087)	(1.039)	(0.48)	(-5.716)	(-2.128)
FTSE4GOOD GLB 100	-0.107***	-0.032	-0.002	-0.003	-0.141***
	(-3.491)	(-0.939)	(-0.271)	(-0.256)	(-4.811)
FTSE4GOOD GLB	-0.107***	0.019	0.001	-0.021*	-0.109***
	(-4.398)	(0.697)	(0.35)	(-1.727)	(-4.059)
FTSE4GOOD JAP	0.016	-0.064***	0.002	0.014**	-0.017
	(0.696)	(-3.052)	(0.349)	(2.506)	(-1.109)
FTSE4GOOD US 100	-0.148***	-0.025	0.003	0.015*	-0.047***
	(-9.037)	(-1.187)	(0.484)	(1.752)	(-2.983)
FTSE4GOOD US	-0.138***	-0.005	0.002	0.008	-0.073***
	(-9.767)	(-0.249)	(0.314)	(1.017)	(-4.549)
J-stat (1)	76.083***	44.285***	3.778	42.959***	43.948***
J-stat (2)					
MSCI EUROPE ESG	-0.033	-0.054***	-0.002	-0.031**	-0.067***
	(-1.338)	(-3.158)	(-0.333)	(-2.485)	(-3.146)
MSCI JAPAN ESG	0	0.001	0.019	0.057*	-0.024
	(0.002)	(0.013)	(0.862)	(1.783)	(-0.377)
MSCI NORTH AMERICA ESG	-0.006	0.137***	0.003	-0.03***	0.005
	(-0.414)	(10.218)	(0.736)	(-4.995)	(0.405)
MSCI WORLD ESG	-0.009	0.098***	0.002	-0.008	-0.024**
	(-0.748)	(6.425)	(0.59)	(-1.383)	(-2.054)
J-stat (1)	1.958	48.838***	1.453	22.869***	8.656*
J-stat (2)					

SRI Index	DC.HML	DC.WML	DC.RMW	DC.CMA
FTSE4GOOD EUR 50	0.161***	0.067***	0.056	-0.135***
	(5.113)	(3.788)	(1.313)	(-3.595)
FTSE4GOOD EUR	0.128***	-0.001	0.064***	-0.058***
	(7.83)	(-0.092)	(3.184)	(-2.976)
FTSE4GOOD GLB 100	0.098***	0.09***	-0.106**	0.08
	(2.686)	(4.821)	(-2.161)	(1.499)
FTSE4GOOD GLB	0.082***	0.026*	-0.053	0.024
	(2.696)	(1.72)	(-1.144)	(0.522)
FTSE4GOOD JAP	-0.059***	-0.022	-0.07**	0.052*
	(-2.932)	(-1.515)	(-2.372)	(1.691)
FTSE4GOOD US 100	-0.088***	0.028**	-0.128***	-0.003
	(-4.757)	(2.577)	(-5.403)	(-0.104)
FTSE4GOOD US	-0.082***	0.025***	-0.124***	-0.036
	(-4.665)	(2.586)	(-5.592)	(-1.402)
J-stat (1)	48.66***	45.538***	40.562***	37.414***
J-stat (2)				
MSCI EUROPE ESG	0.25***	0	0.123*	0.035
	(4.449)	(0.015)	(1.85)	(1.156)
MSCI JAPAN ESG	0.009	-0.02	-0.077	-0.077
	(0.123)	(-0.328)	(-0.668)	(-0.686)
MSCI NORTH AMERICA ESG	0.048***	-0.011	0.006	-0.129***
	(3.154)	(-1.337)	(0.305)	(-5.524)
MSCI WORLD ESG	0.12***	0	0.068***	-0.069**
	(6.676)	(-0.033)	(3.285)	(-2.551)
J-stat (1)	31.201***	4.139	12.565**	29.043***
J-stat (2)				

SRI Index	**IP1**	**IP2**	**IP3**	**Adj. R2**
FTSE4GOOD EUR 50	-0.011***	-0.005*	0.006*	0.985
	(-6.636)	(-1.732)	(1.896)	
FTSE4GOOD EUR	-0.003***	-0.004***	-0.003*	0.994
	(-4.064)	(-3.166)	(-1.717)	
FTSE4GOOD GLB 100	-0.011***	-0.006*	-0.04***	0.962
	(-4.847)	(-1.701)	(-6.728)	
FTSE4GOOD GLB	-0.019***	-0.002	-0.038***	0.971
	(-8.681)	(-0.714)	(-8.523)	
FTSE4GOOD JAP	-0.002*	-0.003*	0.001	0.988
	(-1.725)	(-1.808)	(0.472)	
FTSE4GOOD US 100	0.007***	-0.043***	-0.064***	0.984
	(5.928)	(-15.759)	(-19.234)	
FTSE4GOOD US	0.007***	-0.041***	-0.067***	0.983
	(6.195)	(-15.642)	(-18.984)	
J-stat (1)	49.232***	110.377***	117.086***	
J-stat (2)				
MSCI EUROPE ESG	0.002**	-0.008***	-0.007***	0.993
	(2.025)	(-3.862)	(-2.622)	
MSCI JAPAN ESG	-0.039***	0.021**	-0.012	0.807
	(-6.715)	(2.119)	(-0.994)	
MSCI NORTH AMERICA ESG	-0.002*	-0.001	-0.012***	0.993
	(-1.694)	(-0.36)	(-4.247)	
MSCI WORLD ESG	-0.005***	-0.002	-0.013***	0.995
	(-6.162)	(-1.574)	(-6.229)	
J-stat (1)	65.414***	17.752***	26.865***	
J-stat (2)				

4.2.5 Variance Inflation Factor Analysis

In table 13 we summarise the results of the Variance Inflation Factor Analysis of each Fama-French (2015) regression used in a three GMM System framework. As can be seen, we don't detect problems of multicollinearity in almost any regression except in one case, MSCI ESG Europe. These results are expected since we also didn't detect substantial changes of sign or significance when we added more variables from the more complex models, which would also occur in the presence of such problems (Russo et al., 2016). Moreover, we obtained many estimates that were statistically significant at 1%, which would not be possible in the presence of severe multicollinearity problems.

Table 13 - Variance Inflation Factors of each regressor in each Fama-French (2015) Regression
Source: Adapted from econometric software Eviews 9. We present the Centered Variance Inflation Factor of each regressor used in each regression the three GMM Systems. Ben-rf: difference in return between the conventional official benchmark and risk-free rate. The remaining regressors were already defined. Sample (observations): FTSE4GOOD - 02/12/2004 – 31/05/2017 (3260); MSCI ESG - 01/10/2007 – 31/05/2017 (2523); STOXX ESG - 28/07/2013 – 31/05/2017 (1048).

SRI Index	Ben-rf	SMB	HML	WML	RMW	CMA
FTSE4GOOD EUR 50	2.121341	2.090595	4.285036	1.85102	2.496588	1.754993
FTSE4GOOD EUR	2.285049	1.668989	5.081373	2.154177	2.91335	1.8925
FTSE4GOOD GLB 100	2.105008	1.297322	2.246815	1.771961	1.294412	1.919042
FTSE4GOOD GLB	2.358857	1.282153	1.823799	1.574586	1.388449	1.874491
FTSE4GOOD JAP	1.118177	1.256514	1.473288	1.21798	1.614183	1.667217
FTSE4GOOD US 100	1.858654	1.45638	1.65808	1.230316	1.262305	1.78468
FTSE4GOOD US	1.666911	1.269959	1.703477	1.274763	1.267669	1.870155
MSCI EUROPE ESG	10.95762	4.141055	31.22408	4.513774	16.64022	2.673208
MSCI JAPAN ESG	1.333222	1.439827	1.355835	1.564154	1.253537	1.476975
MSCI NORTH AMERICA ESG	1.659522	1.178669	2.205321	1.09257	1.575108	2.027494
MSCI WORLD ESG	1.862605	1.532762	2.156384	1.489356	1.333542	1.401611
STOXX AP ESG LDRS 50	6.218771	4.913017	2.277548	3.667789	1.792978	2.385151
STOXX EU ESG LDRS 50	1.078711	1.102133	2.329742	1.159322	1.811295	1.652011
STOXX GLOBAL ESG LEADERS	1.684249	1.466898	3.04269	2.624371	1.704753	2.058638
STOXX N.AMR ESG LDRS 50	1.106809	1.019367	2.367023	1.248215	1.516156	2.244612

4.3 Summary of Results

In this subsection, we summarise the results we obtained throughout our analysis complemented by Table 14, in which we provide a more clear and intelligible depiction of the results as a whole in terms of the signal of alpha and other risk-exposures we detected globally and within each Series.

Overall, most alphas are not statistically significant, and most SRI Indices follow the movements of their conventional official benchmarks very closely. However, they still show differences in the return or risk profile in relation to their conventional official benchmarks. This could be seen both by conducting a spanning test, which rejected this hypothesis for most SRI Indices, and by the application of Fama-French based models, in which the addition of risk-factors was statistically significant in almost all series and indices.

The application of such models also resulted in other important findings. One of the most consistent patterns we discovered is that SRI Indices, in general, tend to be biased towards loser stocks. This finding contrasts with the literature in which various patterns were found. The addition of industry factors to Fama-French Five factor model in the analysis of SRI Indices also proved to be relevant. We found the presence of industry effects which is coherent with findings from authors in the literature such as Derwall et al. (2005) and Erragragui and Revelli (2016). Its application caused changes in some estimates but did not absorb the extant biases, implying that tilts of SRI Indices are not justified by industry exposures.

The addition of a Crisis dummy confirmed some findings of the literature and shed light on new facts. There's little evidence of time-varying alphas but most tend to increase during market downturns. Most SRI Indices also appear to be less risky during crisis periods. The most important finding, however, is that most indices showed a statistically significant growth bias during non-crisis periods which shifted into a statistically significant value bias during market downturns. These results show that value companies are more able to keep ethical standards during periods of market turmoil and, therefore, remain included in SRI

Indices while growth companies have to choose survival at the expense of environmental, social and governance standards, being excluded more often as a result.

Throughout our analysis, results are very heterogeneous at various levels. At the country level, we found many differences between indices from different regions which were attested by constructing regional portfolios composed of indices with a similar geographic scope. However, we found that Global SRI Indices have a higher Beta than regional SRI Indices. This result is interesting because of two reasons. First, it contradicts modern finance theory, which predicts that a higher degree of diversification should lead to less exposure to risk, which theoretically should be the case since global Indices are not so restricted in terms of geographic scope as Regional Indices. Second, nonetheless, other authors in the Literature such as Becchetti et al. (2015) and Leite et al. (2014) have also confirmed these patterns by analysing SRI Funds.

We also discovered many differences between indices with similar geographic scope, which suggests differences between the three Series. FTSE4GOOD Indices have a statistically significant large cap bias while MSCI ESG and STOXX ESG Indices tend to, respectively, show no clear pattern and a statistically significant small cap bias. FTSE4GOOD indices also tend to be more invested in companies with weak profitability profiles while indices from other series show different estimates. The large cap bias from FTSE4GOOD Indices is consistent with findings in the literature, but other results represent new facts uncovered by the application of Fama-French models.

These differences were also reflected in discovery of particularly idiosyncratic indices with respect to risk-exposures. The most notorious example is STOXX Global ESG Leaders which showed very high risk-exposures, namely in terms of Beta and exposure to Conservative Stocks.

Some estimates also appear to be gradually affected by the application of more complex models and procedures. One example is Alpha. Despite remaining overall non-statistically significant during the analysis, there was a gradual change of sign from negative to positive as additional models and procedures were applied. Another example are estimates of the value factor HML. In the initial application of the Fama-French Model they were mostly positive, revealing a value bias, but as additional procedures were applied they transformed into a growth bias. Estimates of the CMA also experienced a similar transformation, gradually from overall positive to different patterns within and across the Series. These examples also show the importance of applying different models to avoid draw erroneous conclusions.

Finally, besides being robust to heteroskedasticity and autocorrelation, overall, our results are free of multicollinearity problems.

Table 14 - Summary of Positive Estimates

Source: own elaboration. We summarise the number of positive estimates of each regressor in the various models we employed in our analysis of the FTSE4GOOD, MSCI ESG and STOXX ESG Series. For simplification we only provide the number of positive estimates, since negative estimates also provides the same information.

CAPM						
Series (number of indices)	Alpha					
FTSE4GOOD (7)	0					
MSCI ESG (4)	2					
STOXX ESG (4)	1					
Total (15)	3					
Fama-French (1993)						
Series (number of indices)	Alpha	SMB	HML			
FTSE4GOOD (7)	0	0	7			
MSCI ESG (4)	2	2	3			
STOXX ESG (4)	2	3	4			
Total (15)	4	5	14			
Carhart (1997)						
Series (number of indices)	Alpha	SMB	HML	WML		
FTSE4GOOD (7)	1	0	5	2		
MSCI ESG (4)	2	2	2	2		
STOXX ESG (4)	2	3	3	0		
Total (15)	5	5	10	4		
Fama-French (2015)						
Series (number of indices)	Alpha	SMB	HML	WML	RMW	CMA
FTSE4GOOD (7)	2	0	5	1	2	5
MSCI ESG (4)	1	2	2	2	3	3
STOXX ESG (4)	3	3	4	0	3	3
Total (15)	6	5	11	3	8	11
Industry-Adjusted Regressions						
Series (number of indices)	Alpha	SMB	HML	WML	RMW	CMA
FTSE4GOOD (7)	5	0	3	2	1	3
MSCI ESG (4)	1	2	2	2	2	2
STOXX ESG (4)	2	3	3	0	2	3
Total (15)	8	5	8	4	5	8

Single GMM System							
Series (number of indices)	Alpha	SMB	HML	WML	RMW	CMA	
FTSE4GOOD (7)	7	0	1	1	0	3	
MSCI ESG (4)	1	3	0	2	2	2	
STOXX ESG (4)	2	3	3	0	2	3	
Total (15)	10	6	4	3	4	8	
Regional Portfolios							
	Alpha	SMB	HML	WML	RMW	CMA	
Portfolios (4)	2	2	2	0	2	2	
Crisis Dummy							
	Alpha	SMB	HML	WML	RMW	CMA	
FTSE4GOOD (7)	4	0	1	1	2	3	
MSCI ESG (4)	1	3	0	2	1	3	
Total (11)	5	3	1	3	3	6	
Series	DC	DC.Beta	DC.SMB	DC.HML	DC.WML	DC.RMW	DC.CMA
FTSE4GOOD (7)	5	3	0	4	5	2	3
MSCI ESG (4)	3	1	1	4	2	3	1
Total (11)	8	4	1	8	7	5	4

5. Conclusions and Directions for Future Research

5.1 Conclusions

Our Review of the literature has provided an important contribute to the literature. Not only we conducted an extensive analysis of the extant SRI literature but also analysed other aspects which so far, to the best of our knowledge, were not the subject of intense scrutiny, namely SRI Biases uncovered by the application of Fama-French models. This analysis lead us to important conclusions.

Firstly, most authors have been focused on analysing SRI Funds rather than SRI Indices, which may be a more reliable source of SRI. Secondly, there are very different findings about performance related to the application of different approaches at various levels, such as geographic scope, time-frames, among others. Thirdly, Fama-French models were seldom used in the analysis of SRI Funds, despite the importance of its application, namely in providing a more accurate description of SRI Indices not only in terms of performance but also on what biases they may be subject to. Finally, many authors have taken into account the possibility that performance measures and risk-exposures are time-variant. They found evidence to support the latter, either by applying conditional approaches or comparing crisis with non-crisis periods.

Based on these insights, our empirical work provides two additional contributions. The first is the analysis SRI Indices with diverse regional scope from three different Series, FTSE4GOOD, MSCI ESG and STOXX ESG. Therefore, we analyse an important source of SRI that has received less attention and avoid sample specific results by relying on multiple sources of SRI and focusing on different geographic scopes. The second is the application of procedures which, to the best of our knowledge, were rarely or not applied at all in the analysis of SRI Indices. Besides the application of Fama-French models, we also employ other methods such as the addition of industry factors, the application of a Generalized Method of Moments-system framework and the comparison between crisis and non-crisis periods.

Our results show that, overall, alphas are overall non-statistically significant and most SRI Indices follow very closely the movements of their respective conventional official benchmarks. Nonetheless, there was evidence of differences between them, both by conducting a spanning test and by the addition of Fama-French factors, which were statistically significant both at the series and the index level. The application of these models and other procedures also allowed the detection of important patterns.

SRI Indices tend to be biased towards loser stocks. We found the presence of industry-effects but also that they don't justify other extant biases. Risk-adjusted performances increases during crisis periods but results are not statistically significant. SRI Indices tend to be less risky and more exposed to Value Companies. However, overall there are multiple differences between SRI Indices. As expected, SRI Indices with different geographic scopes show different patterns in terms of risk-exposures. In this respect, an interesting finding is that Global SRI Indices tend to have a higher Beta than regional SRI indices, which, despite being a counterintuitive result that contradicts modern finance theory, is consistent with other findings in the literature.

However, there are also differences between SRI Indices even when controlling for geographic scope, industry effects and period of analysis. Most notably, FTSE4GOOD Indices tend to be more biased towards large companies with weak profitability profiles while indices from other series tend to have either no patterns or a small cap bias. These differences confirm the importance of analysing SRI Indices using different sources. This could possibly be explained by differences in the screening processes of the different agencies responsible for SRI Indices, which may result in the inclusion with companies with different attributes.

Moreover, some SRI Indices show very idiosyncratic behaviour. The most notable example is STOXX Global ESG, which showed both a much higher Beta and exposure to Conservative stocks than other indices. This highlights the importance of conducting individual analysis of SRI Vehicles to avoid missing important features.

Finally, we also found that the results of some estimates are sensitive to the application of additional models and procedures, namely when changing the period of analysis. This occurred with the sign of most alphas, which gradually went from negative to positive, and estimates of the value and investment factor.

5.2 Directions for Future Research

Our results provide several possible avenues for future research to explore. One possibility is to conduct a similar analysis to other well-known SRI Series, such as the Dow Jones Sustainability Indices, Ethibel Indices, among others. This would contribute to confirm whether our findings are representative of SRI in general. It would also be of interest to see the development of more theoretical papers that attempt to explain possible reasons that explain the differences found the biases found in SRI biases. In particular, why do SRI Indices tend to select loser stocks? Moreover, why do value companies tend to be more present in SRI Indices during times of crisis?

Despite the different patterns, we confirmed differences in risk-exposures between indices belonging to the three Series. We confirmed that FTSE4GOOD Indices have a large cap bias and are more invested in companies with weak profitability profiles but also discovered that indices from MSCI ESG and STOXX ESG have, respectively, different patterns and a small cap bias. Future research should focus on determining what factors contribute to these different attributes. This is important to allow a quicker standardization of the screening processes of different SRI providers.

One interesting finding was also the fact that Global SRI Indices had a higher risk than Regional SRI Indices in relation to their conventional official benchmarks. This contradicts finance theory according to which more diversification should to less risk exposure. The fact that our findings are coherent with other authors from the literature such as Becchetti et al. (2015) reinforces the need for further research on this topic.

As we have mentioned, SRI Indices are not subject to transaction costs and managements fees and deliver similar returns as their conventional benchmarks. SRI Funds, on the other hand, are subject to such constraints. However, as we have shown in the Review of the Literature, most papers that study SRI Funds have also found that they perform similarly as their conventional counterparts. These results beg the question: How can SRI Funds deliver similar returns as SRI Indices if they are subject to management fees and transaction costs? Is this accomplished by neglecting ethical obligations? By, for instance, investing in more lucrative but less ethical stocks? Or this accomplished by specific characteristics of funds, such as active management?

Future research should focus on more on the ethical nature of SRI Funds and on conducting more detailed comparisons between SRI Funds and Indices. An example would be conducting a similar analysis to Funds that tracked the indices from the analysed series and compare their findings to those of this Thesis.

References

Abdymomunov, A. and J. Morley (2011), "Time Variation of CAPM Betas across Market Volatility Regimes", *Applied Financial Economics,* Vol. 21, N. 19, pp. 1463-1478.

Amenc, N. and V. Le Sourd (2010), "The Performance of Socially Responsible Investment and Sustainable Development in France: An Update after the Financial Crisis". EDHEC-Risk Institute.

Andersen, M. L. and N. Dakota (2003), "Corporate Social and Financial Performance: A Meta-Analysis", *Organization Studies,* Vol. 15, N. 2, pp. 17-38.

Areal, N., M. C. Cortez and F. Silva (2013), "The Conditional Performance of US Mutual Funds over Different Market Regimes: Do Different Types of Ethical Screens Matter?", *Financial Markets and Portfolio Management,* Vol. 27, N. 4, pp. 397-429.

Asness, C., A. Frazzini and L. H. Pedersen (2013), "Quality Minus Junk". University and AQR Capital Management, New York Working Paper N. 1.

Auer, B. R. and F. Schuhmacher (2016), "Do Socially (Ir)responsible Investments pay? New Evidence from International ESG data", *Quarterly Review of Economics and Finance,* Vol. 59, N. 1, pp. 51-62.

Barnett, M. L. and R. M. Salomon (2006), "Beyond Dichotomoy: The Curvilinear Relationship between Social Responsability and Financial Performance", *Strategic Management Journal,* Vol. 27, N. 1, pp. 1101-1122.

Bauer, R., J. Derwall and R. Otten (2007), "The Ethical Mutual Fund Performance Debate: New Evidence from Canada", *Journal of Business Ethics,* Vol. 70, N. 2, pp. 111-124.

Bauer, R., K. Koedijk and R. Otten (2005), "International Evidence on Ethical Mutual Fund Performance and Investment Style", *Journal of Banking and Finance,* Vol. 29, N. 7, pp. 1751-1767.

Bauer, R., R. Otten and A. T. Rad (2006), "Ethical investing in Australia: Is there a Financial Penalty?", *Pacific-Basin Finance Journal*, Vol. 14, N. 1, pp. 33-48.

Becchetti, L., R. Ciciretti, A. Dalò and S. Herzel (2015), "Socially Responsible and Conventional Investment Funds: Performance Comparison and the Global Financial Crisis", *Applied Economics*, Vol. 47, N. 25, pp. 2541-2562.

Belghitar, Y., E. Clark and N. Deshmukh (2014), "Does it pay to be Ethical? Evidence from the FTSE4Good", *Journal of Banking and Finance,* Vol. 47, N. 1, pp. 54-62.

Bello, Z. Y. (2005), "Socially Responsible Investing and Portfolio Diversification", *Journal of Financial Research,* Vol. 28, N. 1, pp. 41-57.

Benson, K. L., T. J. Brailsford and J. E. Humphrey (2013), "Do Socially Responsible Fund Managers Really Invest Differently?", *Journal of Business Ethics,* Vol. 65, N. 4, pp. 337-357.

Blanchett, D. M. (2010), "Exploring the Cost of Investing in Socially Responsible Mutual Funds: An Empirical Study", *Journal of Investing,* Vol. 19, N. 3, pp. 93-103.

Bollen, N. P. B. and J. A. Busse (2001), "On the Timing Ability of Mutual Fund Managers", *Journal of Banking and Finance,* Vol. 37, N. 9, pp. 3654-3668.

Borgers, A., J. Derwall, K. Koedijk and J. ter Horst (2015), "Do Social Factors influence Investment Behavior and Performance? Evidence from Mutual Fund Holdings", *Journal of Banking & Finance,* Vol. 60, N. 1, pp. 112-126.

Brammer, S., C. Brooks and S. Pavelin (2006), "Corporate Social Performance and Stock Returns: UK Evidence from Disagggregate Measures", *Financial Management,* Vol. 35, N. 3, pp. 97-116.

Brammer, S., C. Brooks and S. Pavelin (2009), "The Stock Performance of America's 100 Best Corporate Citizens", *Quarterly Review of Economics and Finance,* Vol. 49, N. 3, pp. 1065-1080.

Brooks, R. D., R. W. Faff and M. D. McKenzie (1998), "Time-Varying Beta Risk of Australian Industry Portfolios: A Comparison of Modelling Techniques", *Australian Journal of Management,* Vol. 23, N. 1, pp. 1-22.

Capelle-Blancard, G. and S. Monjon (2014), "The Performance of Socially Responsible Funds: Does the Screening Process Matter?", *European Financial Management,* Vol. 20, N. 3, pp. 494-520.

Carhart, M. M. (1997), "On Persistence in Mutual Fund Performance", *Journal of Finance,* Vol. 52, N. 1, pp. 57-82.

Charfeddine, L., A. Najah and F. Teulon (2016), "Socially Responsible Investing and Islamic funds: New Perspectives for Portfolio Allocation", *Research in International Business and Finance,* Vol. 36, N. 1, pp. 351-361.

Christopherson, J. A., W. E. Ferson and D. A. Glassman (1998), "Conditioning Manager Alphas on Economic Information: Another Look at the Persistence of Performance", *Review of Financial Studies,* Vol. 11, N. 1, pp. 111-142.

Climent, F. and P. Soriano (2011), "Green and Good? The Investment Performance of US Environmental Mutual Funds", *Journal of Business Ethics,* Vol. 103, N. 2, pp. 275-287.

Collison, D. J., G. Cobb, D. M. Power and L. A. Stevenson (2008), "The Financial Performance of the FTSE4Good indices", *Corporate Social Responsibility and Environmental Management,* Vol. 15, N. 1, pp. 14-28.

Consolandi, C., A. Jaiswal-Dale, E. Poggiani and A. Vercelli (2009), "Global Standards and Ethical Stock Indexes: The Case of the Dow Jones Sustainability Stoxx Index", *Journal of Business Ethics,* Vol. 87, N. 1 (Supplement), pp. 185-197.

Cortez, M. C., F. Silva and N. Areal (2009), "The Performance of European Socially Responsible Funds", *Journal of Business Ethics,* Vol. 87, N. 4, pp. 573-588.

Cortez, M. C., F. Silva and N. Areal (2012), "Socially Responsible Investing in the Global Market: The Performance of US and European Funds", *International Journal of Finance and Economics,* Vol. 17, N. 3, pp. 254-271.

Cummings, L. S. (2000), "The Financial Performance of Ethical Investment Trusts: An Australian Perspective", *Journal of Business Ethics*, Vol. 25, N. 1, pp. 79-92.

Dam, L. and B. Scholtens (2015), "Toward A Theory Of Responsible Investing: On The Economic Foundations Of Corporate Social Responsibility", *Resource and Energy Economics,* Vol. 41, N. 1, pp. 103-121.

Derwall, J., N. Guenster, R. Bauer and K. Koedijk (2005), "The Eco-Efficiency Premium Puzzle", *Financial Analysts Journal,* Vol. 61, N. 2, pp. 51-63.

Derwall, J. and K. Koedijk (2009), "Socially Responsible Fixed-Income Funds", *Journal of Business Finance and Accounting,* Vol. 36, N. 1-2, pp. 210-229.

Diltz, J. D. (1995), "Does Social Screening Affect Portfolio Performance?", *The Journal of Investing,* Vol. 4, N. 1, pp. 64-69.

Dravenstott, J. and N. Chieffe (2011), "Corporate Social Responsibility: Should I Invest for It or against It?", *The Journal of Investing*, Vol. 20, N. 3, pp. 108-117.

Duuren, E. V., A. Plantinga and B. Scholtens (2016), "ESG Integration and the Investment Management Process: Fundamental Investing Reinvented", *Journal of Business Ethics,* Vol. 138, N. 3, pp. 525-533.

Edmans, A. (2011), "Does the Stock Market fully Value Intangibles? Employee Satisfaction and Equity Prices", *Journal of Financial Economics,* Vol. 101, N. 3, pp. 621-640.

Elton, E. J. and M. J. Gruber (1993), "The Performance of Bond Mutual Funds", *The Journal of Business,* Vol. 66, N. 3, pp. 371-403.

Erragragui, E. and C. Revelli (2016), "Is It Costly To Be Both Shariah Compliant And Socially Responsible?", *Review of Financial Economics*, Vol. 31, N. 11, pp. 64-74.

Escrig-Olmedo, E., M. J. Munoz-Torres and M. A. Fernandez-Izquierdo (2010), "Socially Responsible Investing: Sustainability Indices, ESG Rating and Information Provider Agencies", *International Journal of Sustainable Economy,* Vol. 2, N. 4, pp. 442-461.

Faff, R. W., D. Hiller and J. Hiller (2000), "Time-Varying Beta Risk: An Analysis of Alternative Modelling Techniques", *Journal of Business Finance & Accounting,* Vol. 27, N. 5-6, pp. 523-554.

Fama, E. and K. French (1993), "Common Risk Factors In The Returns On Stocks And Bonds", *Journal of Financial Economics,* Vol. 33, N. 1, pp. 3-56.

Fama, E. and K. French (2015), "A Five-Factor Asset Pricing Model", *Journal of Financial Economics,* Vol. 116, N. 1, pp. 1-22.

Fama, E. and J. D. MacBeth (1974), "Risk, Return and Equilibrium: Empirical Tests", *Journal of Political Economy,* Vol. 81, N. 3, pp. 607-636.

Fernandez-Izquierdo, A. and J. C. Matallin-Saez (2008), "Performance of Ethical Mutual Funds in Spain: Sacrifice or premium?", *Journal of Business Ethics,* Vol. 81, N. 2, pp. 247-260.

Ferson, W., T. Henry and D. Kisgen (2006), "Evaluating Fixed Income Fund Returns With Stochastic Discount Factors", *Review of Financial Studies,* Vol. 19, N. 2, pp. 423-455.

Ferson, W. and R. W. Schadt (1996), "Measuring Fund Strategy and Performance in Changing Economic Conditions", *The Journal of Finance,* Vol. 51, N. 2, pp. 425-425.

Filbeck, G., R. Gorman and X. Zhao (2009), "The "Best Corporate Citizens": Are They Good for Their Shareholders?", *Financial Review,* Vol. 44, N. 2, pp. 239-262.

Fowler, S. J. and C. Hope (2007), "A Critical Review Of Sustainable Business Indices And Their Impact", *Journal of Business Ethics,* Vol. 76, N. 3, pp. 243-252.

Galema, R., A. Plantinga and B. Scholtens (2008), "The Stocks At Stake: Return And Risk In Socially Responsible Investment", *Journal of Banking and Finance,* Vol. 32, N. 12, pp. 2646-2654.

Gil-Bazo, J., P. Ruiz-Verdú and A. A. P. Santos (2010), "The Performance Of Socially Responsible Mutual Funds: The Role Of Fees And Management Companies", *Journal of Business Ethics,* Vol. 94, N. 2, pp. 243-263.

Girard, E., H. Rahman and B. Stone (2007), "Socially Responsible Investments: Goody-Two-Shoes or Bad to the Bone?", *Journal of Investing,* Vol. 16, N. 1, pp. 96-110.

Goldreyer, E. F. and J. D. Diltz (1999), "The Performance Of Socially Responsible Mutual Funds: Incorporating Sociopolitical Information In Portfolio Selection", *Managerial Finance*, Vol. 25, N. 1, pp. 23-36.

Gregory, A., J. Matatko and R. Luther (1997), "Ethical Unit Trust Financial Performance: Small Company Effects And Fund Size Effects", *Journal of Business Finance and Accounting,* Vol. 24, N. 5, pp. 705-725.

Guerard, J. B. (1997), "Is There a Cost to Being Socially Responsible in Investing?", *The Journal of Investing,* Vol. 6, N. 2, pp. 11-18.

Gujarati, D. N. and C. P. Dawn (2009), *Basic Econometrics*, New York: McGraw-Hill Irwin.

Hamilton, S., H. Jo and M. Statman (1993), "Doing Well While Doing Good? The Investment Performance of Socially Responsible Mutual Funds", *Financial Analysts Journal,* Vol. 49, N. 6, pp. 62-66.

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica,* Vol. 50, N. 4, pp. 1029-1054.

Hayashi, F. (2000), *Econometrics,* Princeton: Princeton Univ. Press.

Henrikson, R. and R. Merton (1981), "On Market Timing and Investment Performance II. Statistical Procedures for Evaluating Forecasting Skills", *Journal of Business,* Vol. 54, N. 4, pp. 513-533.

Hill, R. P., T. Ainscough, T. Shank and D. Manullang (2007), "Corporate Social Responsibility And Socially Responsible Investing: A Global Perspective", *Journal of Business Ethics*, Vol. 70, N. 2, pp. 165-174.

Holmes, K. A. and R. W. Faff (2004), "Stability, Asymmetry And Seasonality of Australian Multi-Sector Managed Funds within a Dynamic Kalman Filter Framework", *International Review of Financial Analysis,* Vol. 17, N. 5, pp. 998-1011.

Hooi, H. L., R. A. Wei and R. Smyth (2015), "Performance and Performance Persistence of Socially Responsible Investment Funds in Europe and North America", *The North American Journal of Economics and Finance,* Vol. 34, N. 1, pp. 254-266.

Hou, K., C. Xue and L. Zhang (2015), "Digesting Anomalies: An Investment Approach", *Review of Financial Studies,* Vol. 28, N. 3, pp. 650-705.

Humphrey, J. E., D. D. Lee and Y. Shen (2012), "Does It Cost To Be Sustainable?", *Journal of Corporate Finance,* Vol. 18, N. 3, pp. 626-639.

Ito, Y., S. Managi and A. Matsuda (2012), "Performance of Socially Responsible Investment and Environmentally Friendly Funds", *Journal of the Operational Research Society,* Vol. 64, N. 11, pp. 1583-1594.

Jegadeesh, N. and S. Titman (1993), "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency", *The Journal of Finance,* Vol. 48, N. 1, pp. 65-91.

Jensen, M. (1968), "The Performance of Mutual Funds in the Period 1945-1964", *Journal of Finance*, Vol. 23, N. 2, pp. 389-416.

Jin, H. H., O. S. Mitchell and J. Piggott (2006), "Socially Responsible Investment In Japanese Pensions", *Pacific Basin Finance Journal,* Vol. 14, N. 5, pp. 427-438.

Johnson, R. A. and D. W. Wichtern (2007), *Applied Multivariate Statistical Analysis*, New Jersey: Pearson.

Jones, S., S. Van Der Laan, G. Frost and J. Loftus (2008), "The Investment Performance Of Socially Responsible Investment Funds In Australia", *Journal of Business Ethics,* Vol. 80, N. 2, pp. 181-203.

Kempf, A. and P. Osthoff (2007), "The Effect Of Socially Responsible Investing On Portfolio Performance", *European Financial Management,* Vol. 13, N. 5, pp. 908-922.

Kreander, N., R. H. Gray, D. M. Power and C. D. Sinclair (2005), "Evaluating The Performance Of Ethical And Non-Ethical Funds: A Matched Pair Analysis", *Journal of Business Finance and Accounting,* Vol. 32, N. 7-8, pp. 1465-1493.

Kurtz, L. and DiBartolomeo (2011), "The Long-Term Performance of a Social Investment Universe", *The Journal of Investing,* Vol. 20, N. 3, pp. 95-102.

Kurtz, L. and D. DiBartolomeo (1999), "Managing Risk Exposures of Socially Screened Portfolios", *Technical Report, Northfield Information Services, Boston, MA,* Vol. 1, N. 415, pp. 1-17.

Lee, D. D. and R. W. Faff (2009), "Corporate Sustainability Performance and Idiosyncratic Risk: A Global Perspective", *The Financial Review,* Vol. 44, N. 1, pp. 213-237.

Lee, D. D., J. E. Humphrey, K. L. Benson and J. Y. K. Ahn (2010), "Socially Responsible Investment Fund Performance: The Impact Of Screening Intensity", *Accounting & Finance,* Vol. 50, N. 2, pp. 351-370.

Leite, P. and M. C. Cortez (2014), "Selectivity And Timing Abilities Of International Socially Responsible Funds", *Applied Economics Letters,* Vol. 21, N. 3, pp. 185-188.

Leite, P., M. C. Cortez and M. Céu Cortez (2015), "Performance Of European Socially Responsible Funds During Market Crises: Evidence From France", *International Review of Financial Analysis,* Vol. 40, N. 1, pp. 132-141.

Lesser, K., S. Lobe and C. Walkshäusl (2014), "Green And Socially Responsible Investing In International Markets", *Journal of Asset Management,* Vol. 15, N. 5, pp. 317-331.

Lesser, K., F. Rößle and C. Walkshäusl (2016), "Socially Responsible, Green, And Faith-Based Investment Strategies: Screening Activity Matters!", *Finance Research Letters,* Vol. 16, N. 2, pp. 171-178.

Lin, C.-H., H.-L. Yang and D.-Y. Liou (2009), "The Impact of Corporate Social Responsibility on Financial Performance: Evidence from Business in Taiwan", *Technology in Society,* Vol. 31, N. 1, pp. 56-63.

Lins, K. V., H. Servaes and T. Ane (2016), "Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis", *The Journal of Finance,* Vol. 72, N. 4, pp. 1785-1824.

Luther, R. G., J. Matatko and D. C. Corner (1992), "The Investment Performance of UK “Ethical” Unit Trusts", *Accounting, Auditing & Accountability Journal,* Vol. 5, N. 4, pp. 57-70.

Mallin, C. A., B. Saadouni and R. J. Briston (1995), "The Financial Performance Of Ethical Investment Funds", *Journal of Business Finance & Accounting,* Vol. 22, N. 4, pp. 483-496.

Managi, S. (2012), "Do Socially Responsible Investment Indexes Outperform Conventional Indexes?", *Applied Financial Economics,* Vol. 22, N. 4, pp. 37-41.

Margolis, J. and J. P. Walsh (2001), *People and Profits? The Search for a Link Between a Company's Social and Financial Performance*, New York: Psychology Press.

Margolis, J. D. and J. P. Walsh (2003), "Misery Loves Companies: Rethinking Social Initiatives by Business", *Administrative Science Quarterly,* Vol. 48, N. 2, pp. 268-268.

Martí-Ballester, C.-P. (2015), "Can Socially Responsible Investment For Cleaner Production Improve The Financial Performance Of Spanish Pension Plans?", *Journal of Cleaner Production*, Vol. 106, N. 1, pp. 466-477.

Mollet, J. C. and A. Ziegler (2014), "Socially Responsible Investing And Stock Performance: New Empirical Evidence For The US And European Stock Markets", *Review of Financial Economics,* Vol. 23, N. 4, pp. 208-216.

Muñoz, F., M. Vargas and I. Marco (2014), "Environmental Mutual Funds: Financial Performance And Managerial Abilities", *Journal of Business Ethics,* Vol. 124, N. 4, pp. 551-569.

Newey, W. K. and K. D. West (1987), "Hypothesis Testing with Efficient Method of Moments Estimation", *International Economic Review,* Vol. 28, N. 3, pp. 777-787.

Nofsinger, J. and A. Varma (2014), "Socially Responsible Funds And Market Crises", *Journal of Banking & Finance,* Vol. 48, N. 11, pp. 180-193.

Ooi, E. and P. Lajbcygier (2013), "Virtue Remains After Removing Sin: Finding Skill Amongst Socially Responsible Investment Managers", *Journal of Business Ethics,* Vol. 113, N. 2, pp. 199-224.

Orlitzky, M., F. L. Schmidt and S. L. Rynes (2003), "Corporate Social and Financial Performance: A Meta-Analysis", *Organization Studies,* Vol. 24, N. 3, pp. 403-441.

Ortas, E., R. L. Burritt and J. M. Moneva (2013), "Socially Responsible Investment And Cleaner Production In The Asia Pacific: Does It Pay To Be Good?", *Journal of Cleaner Production,* Vol. 52, N. 1, pp. 272-280.

Ortas, E., J. M. Moneva and M. Salvador (2012), "Does Socially Responsible Investment Equity Indexes In Emerging Markets Pay Off? Evidence From Brazil", *Emerging Markets Review,* Vol. 13, N. 4, pp. 581-597.

Rehman, R. U., J. Zhang, J. Uppal, C. Cullinan and M. Akram Naseem (2016), "Are Environmental Social Governance Equity Indices A Better Choice For Investors? An Asian Perspective", *Business Ethics: A European Review,* Vol. 25, N. 4, pp. 440-459.

Renneboog, L., J. Ter Horst and C. Zhang (2008), "The Price Of Ethics And Stakeholder Governance: The Performance Of Socially Responsible Mutual Funds", *Journal of Corporate Finance,* Vol. 14, N. 3, pp. 302-322.

Russo, A., M. Mariani and F. Perrini (2016), "Cherry Picking or Depth-Oriented Strategic Investing? Evidence from SRI Activity", *International Journal of Business and Management,* Vol. 11, N. 11, pp. 13-13.

Sauer, D. A. (1997), "The Impact Of Social-Responsibility Screens On Investment Performance: Evidence From The Domini 400 Social Index And Domini Equity Mutual Fund", *Review of Financial Economics,* Vol. 6, N. 2, pp. 137-149.

Scholtens, B. (2005), "Style and Performance of Dutch Socially Responsible Investment Funds", *The Journal of Investing,* Vol. 14, N. 1, pp. 63-72.

Scholtens, B. (2007), "Financial and Social Performance of Socially Responsible Investments in the Netherlands", *Corporate Governance,* Vol. 15, N. 6, pp. 1090-1105.

Schröder, M. (2004), "The Performance Of Socially Responsible Investments: Investment Funds And Indices", *Financial Markets and Portfolio Management,* Vol. 18, N. 2, pp. 122-142.

Schröder, M. (2007), "Is There A Difference? The Performance Characteristics Of SRI Equity Indices", *Journal of Business Finance and Accounting,* Vol. 34, N. 1-2, pp. 331-348.

Schwartz, M. S. (2003), "The "Ethics" of Ethical Investing", *Journal of Business Ethics,* Vol. 43, N. 1, pp. 195-213.

Scowcroft, A. and J. Sefton (2005), "Understanding Momentum", *Financial Analysts Journal,* Vol. 61, N. 2, pp. 61-82.

Silva, F. and M. C. Cortez (2016), "The Performance Of US And European Green Funds In Different Market Conditions", *Journal of Cleaner Production,* Vol. 135, N. 1, pp. 558-566.

Soler-Domínguez, A. and J. C. Matallín-Sáez (2016), "Socially (Ir)Responsible Investing? The Performance Of The VICEX Fund From A Business Cycle Perspective", *Finance Research Letters,* Vol. 16, N. 1, pp. 190-195.

Statman, M. (2000), "Socially Responsible Mutual Funds", *Financial Analysts Journal,* Vol. 56, N. 3, pp. 30-39.

Statman, M. (2006), "Socially Responsible Indexes", *The Journal of Portfolio Management*, Vol. 32, N. 3, pp. 100-109.

Sun, M., K. Nagata and H. Onoda (2011), "The Investigation Of The Current Status Of Socially Responsible Investment Indices", *Journal of Economics and International Finance*, Vol. 3, N. 13, pp. 676-684.

Treynor, J. and K. Mazuy (1966), "Can Mutual Funds Outguess The Market", *Harvard Business Review,* Vol. 44, N. 4, pp. 131-136.

Trinks, P. J. and B. Scholtens (2015), "The Opportunity Cost of Negative Screening in Socially Responsible Investing", *Journal of Business Ethics,* Vol. 140, N. 2, pp. 1-16.

Vermeir, W., E. Van de Velde and F. Corten (2005), "Sustainable and Responsible Performance", *The Journal of Investing,* Vol. 14, N. 3, pp. 94-101.

Wallis, M. V. and C. Klein (2015), "Ethical requirement and financial interest: a literature review on socially responsible investing", *Business Research,* Vol. 8, N. 1, pp. 61-98.

Wimmer, M. (2013), "ESG-persistence in Socially Responsible Mutual Funds", *Journal of Management and Sustainability,* Vol. 3, N. 1, pp. 9-15.

Wu, J., G. Lodorfos, A. Dean and G. Gioulmpaxiotis (2017), "The Market Performance of Socially Responsible Investment during Periods of the Economic Cycle - Illustrated Using the Case of FTSE", *Managerial and Decision Economics,* Vol. 38, N. 2, pp. 238-251.

Xiao, Y., R. Faff, P. Gharghori and B.-K. Min (2015), "The Financial Performance of Socially Responsible Investments: Insights from the Intertemporal CAPM", *Journal of Business Ethics,* Vol. 1, N. 1, pp. 1-12.

Appendices

Appendix 1 – Tables of Review of the Literature

Table 15 - Individual findings of 41 papers at the fund level

Source: own Elaboration. Methodology shows the main models, regressions, performance measures and other approaches used by the authors. Modifications/Variables show, respectively, modifications made to the base models and the explanatory variables used in the return regressions. Period refers to the to the longest time period covered by a study. Countries mention the individual countries which were analysed, except for studies that cover many different ones, in which case we refer to regions instead (for e.g. Europe, North America, etc.). We also refer if studies analysed aggregate (SRI) or individual dimensions. Performance difference typically refers to the difference in risk-adjusted performance between SRI Vehicles and Conventional Investments. Size, Value and Momentum refer to biases detected using Fama-French based models (Fama-French, 1993 and Carhart, 1997) or other approaches. Despite using Fama French based models, some authors don't report risk factor loadings. For sake of simplicity we only considerer three categories of results: positive and statistically significant; not-statistically significant; negative and statistically significant. Other abbreviations: F-F (Fama-French); RAP (Risk-adjusted performance); ER (Excess return); AR (Absolute Return); MRI (Morally Responsible Investments); SRI (traditional SRI Investments); misc. (miscellaneous); n.s.s. (non-statistically significant).

Author	Methodology	Modifications /Variables	Period	Countries	Dimensions	Performance difference	Size	Value	Momentum
Amenc et al. (2010)	F-F (1993) (1)	(1) oil price variations	2002-2009	France	SRI, Environment	n.s.s.	small	dimensions: no pattern (sri), growth (environment)	-
Areal et al. (2013)	CAPM (1) Carhart (1997) (1)	(1) time-varying beta: conditional approach (Abdymonunov and Morley, 2011)	1993-2009	US	SRI, MRI	subperiods: positive (high volatility), negative (low volatility)	small	value	negative
Barnett and Solomon (2006)	CAPM, F-F (1993), Carhart (1997) RAP Regression (1)	(1) screening intensity; fund characteristics ; types of screening	1972-2000	World	Environment, Community, Labor	dimensions: positive (community), negative (environment, labor), curvilinear (intensity)	-	-	-

Author	Methodology	Modifications /Variables	Period	Countries	Dimensions	Performance difference	Size	Value	Momentum
Bauer et al. (2005)	Carhart (1997)	-	1990-2001	World, US, UK, Germany	SRI	subperiods: negative (1990-1993), mixed (1994-1997), non-negative (1998-2001); n.s.s. (1990-2001)	countries: small (UK, Germany), large (US)	growth	no pattern
Bauer et al. (2006)	Carhart (1997) (1,2)	(1) local (2) time-varying beta: conditional approach (Ferson and Schadt, 1996)	1992-2003	Australia	SRI	subperiods: negative (1992-1996); n.s.s. (1992-2003, 1996-2003)	n.s.s.	n.s.s.	n.s.s.
Bauer et al. (2007)	CAPM, Carhart (1997), Sharpe Ratio	-	1994-2003	Canada	SRI	n.s.s.	n.s.s.	n.s.s.	n.s.s.
Becchetti et al. (2015)	CAPM (1), Carhart (1997) (1), Sharpe (2)	(1) Timing (Bollen and Busse, 2001)	1992-2012	World, US, EU, Asia	SRI	subperiods: negative (dot com crisis) positive (financial crisis)	Investment size: small (Large) Large (small) countries: small (EU, Asia) n.s.s. (World) large (US)	value	n.s.s.
Bello (2005)	CAPM, Sharpe ratio, eSDAR, statistical tests	-	1994-2001	US	SRI	n.s.s.	-	-	-

Author	Methodology	Modifications /Variables	Period	Countries	Dimensions	Performance difference	Size	Value	Momentum
Benson et al. (2013)	AR Regression (1,2), Sharpe Ratio, statistical tests	(1) Risk free rate (2) Industry adjusted: returns	1994-2003	US	SRI, religious	n.s.s.	-	-	-
Borgers et al. (2015)	Carhart (1997)	-	2004-2012	US	faith	negative	-	-	-
Capelle-Blanchard and Monjon (2014)	CAPM, F-F (1993), Carhart (1997), RAP Regressions (1)	(1) screening intensity, rating, ESG dimension, fund attributes, investment style	2004-2007	France	ESG, Environment, Social, Governance	intensity: curvilinear	-	-	-
Climent and Soriano (2011)	CAPM, Carhart (1997)	-	1987-2009	US	SRI, Green	subperiods: negative (1987-2001), n.s.s. (2001-2009)	small	n.s.s.	n.s.s.
Cortez et al. (2012)	CAPM (1,2) F-F (1993) (1,2) Carhart (1997) (1,2)	(1) time-varying alpha and beta (Christophers on et al., 1998) (2) local	1996-2008	US, UK, EU	SRI	countries: n.s.s. (EU), negative (US, Austria)	small	value	no pattern
Cummings (2000)	CAPM, Multi-index, Sharpe, Treynor Ratios	-	1986-1996	Australia	SRI	n.s.s.	no pattern	-	-

Author	Methodology	Modifications /Variables	Period	Countries	Dimensions	Performance difference	Size	Value	Momentum
Derwall and Koedijk (2009)	CAPM (1) CAPM (1,2,3) RAP Regressions (4), Sharpe Ratio	(1) Default, Option, Equity (2) term structure (3) Macroeconomic expectation (Elten et al., 1995) (4) size, expenses, turnover, SRI dummy	1987-2003	US	SRI	equity/bond: positive (balanced) equal (bond)	-	-	-
Fernandez-Isquierdo and Matallin-Saez (2008)	Multi-index model	-	1998-2001	Spain	SRI	n.s.s.	-	-	-
Gil-Bazo et al. (2010)	Carhart (1997)	-	1997-2005	US	SRI	fund characteristics: positive (specialized) negative (not specialized)	-	-	-
Girard et al. (2007)	CAPM, CAPM (1)	(1) Timing (Treynor and Mazuy, 1966)	1984-2003	US	SRI	positive	-	-	-
Goldreyer (1999)	CAPM, Sharpe, Treynor	-	1981-1997	US	SRI	strategies/bond : no pattern (inclusion, bond)	-	-	-
Gregory et al. (1997)	CAPM (1)	(1) size index	1986-1994	UK	SRI	n.s.s.	small	-	-
Hamilton (1993)	CAPM	-	1981-1990	US	SRI	n.s.s.	-	-	-

Author	Methodology	Modifications /Variables	Period	Countries	Dimensions	Performance difference	Size	Value	Momentum
Hooi et al. (2015)	F-F (1993), Carhart (1997) (1)	(1) crisis: dummy	2001-2011	US, EU	SRI	positive	countries: small (EU), large (US)	countries: Value (US), growth (EU)	countries: positive (US), negative (EU)
Ito et al. (2012)	CAPM, non-parametric analysis, DMV analysis	-	2000-2009	US, EU	SRI, Environment	positive	-	-	-
Jin et al. (2006)	CAPM (1,2)	(1) SRI, Beta, Fundamentals (2) Industry Adjusted: Dummys	1997-2005	Japan	SRI	misc.: no pattern (pre and post launch, CAPM, benchmark)	large	n.s.s.	-
Jones et al. (2008)	CAPM, Carhart (1997)	-	1986-2005	Australia	SRI	negative	n.s.s.	n.s.s.	n.s.s.
Kreander et al. (2005)	CAPM (1,2) RAP Regression (3), Sharpe, Treynor Ratios	(1) size index (Gregory et al., 1997); (2) timing (Henriksson and Merton, 1981) (3) fund characteristics , dummy: ethical	1995-2001	EU	SRI	n.s.s.	n.s.s.	-	-
Lee et al. (2010)	CAPM, Carhart (1997), Performance Measures Regressions (1), Total Returns, Sharp, Information Ratios, M2	(1) fund characteristics	1986-2006	US	SRI	misc.: negative (intensity, Carhart alpha)	large	value	positive

Author	Methodology	Modifications /Variables	Period	Countries	Dimensions	Performance difference	Size	Value	Momentum
Leite et al. (2014a)	Carhart (1997) (1,2,3,4)	(1) time-varying beta: conditional approach (Ferson and Schaft, 1996) (2) timing (Henriksson and Merton, 1981) (3) market timing (Treynor and Mazuy, 1966) (4) local	2000-2008	EU, UK	SRI	n.s.s.	-	-	-
Leite et al. (2014)	Carhart (1,2)	(1) Time-varying alpha and beta: conditional approach (Christophers on et al., 1998); (2) local	2000-2008	EU, UK	SRI	n.s.s.	n.s.s.	n.s.s.	n.s.s.
Leite et al. (2015)	Carhart (1997) (1,2)	(1) local (2) crisis: dummy	2001-2012	France	SRI	subperiods: n.s.s. (crisis) negative (non-crisis)	subperiods: n.s.s. (crisis) large (non-crisis)	subperiod: growth (non-crisis) n.s.s. (crisis)	subperiod: negative (crisis) n.s.s. (non-crisis)
Lesser et al. (2016)	CAPM, F-F (1993) (1) Carhart (1997) (2)	(1) q-theory (Hou et al., 2015) (2) quality (Asness et al., 2013)	2000-2012	World	Environment, Governance, Social, Broad Green, Energy, Water, Islam, Other	n.s.s.	-	-	-
Luther et al. (1992)	CAPM	-	1972-1990	UK	SRI	-	small	-	-

Author	Methodology	Modifications /Variables	Period	Countries	Dimensions	Performance difference	Size	Value	Momentum
Martí-Ballester (2014)	Multi-index model	-	2007-2013	Spain	Environment, Solidarity	n.s.s.	-	-	-
Munõz et al. (2014)	Carhart (1997)	(1) Timing (Treynor and Mazuy, 1966)	1994-2013	US, EU	SRI, Environment	subperiods: n.s.s. (crisis) negative (non-crisis)	-	-	-
Nofsinger and Varma (2014)	CAPM (1), F-F (1993) (1), Carhart (1997) (1) ER Regressions (2)	(1) crisis: dummy (2) SRI, Fundamentals	2000-2011	US	Environment, Social, Governance, Shareholder Advocacy, Faith	subperiods: positive (crisis), negative (non-crisis)	dimensions: large (SRI); not reported (dimensions)	dimensions: n.s.s. (SRI) not reported (dimensions)	dimensions: negative (SRI) not reported (dimensions)
Ooi and Lajbcygier (2013)	F-F (1993) (1)	(1) filtered benchmarks (no sin industries)	1984-2006	US	SRI	funds: positive (some funds) n.s.s. (others)	n.s.s.	value	-
Renneboog et al. (2008)	CAPM (1) Carhart (1997) (1,2) RAP Regressions (3) Risk Loading Regressions (3,4)	(1) ethical benchmark (2) time-varying beta: conditional approach (Ferson and Schadt, 1996) (3) screening, fund characteristics , fixed effects (4) investment styles	1991-2003	US, Canada, EU, UK, Asia-Pacific.	SRI	countries: n.s.s. (most countries)	Countries: small (Germany, UK) Large (US, Canada, Japan) Risk Loading Regressions: small (in-house research team, ethical screens, older funds) large (shareholder activism)	Countries: Value (Canada, Japan, Norway) Risk Loading Regressions: growth (islam, sin screens)	Countries: n.s.s.; Risk Loading Regressions: negative (high management fees)

Author	Methodology	Modifications /Variables	Period	Countries	Dimensions	Performance difference	Size	Value	Momentum
Russo et al. (2016)	AR Regression (1)	(1) size, screening characteristics, control variables	2000-2005	US	Environment, Governance, Social, Product	dimensions: positive (specific criteria, depth)	-	-	-
Scholtens (2005)	CAPM; Carhart (1997)	-	2001-2003	Netherlands	SRI	n.s.s.	small	value	n.s.s.
Silva and Cortez (2016)	Carhart (1997) (1)	(1) time-varying alpha and beta: conditional approach (Christophers on et al., 1998), Dummy: State of the economy (Ferson and Qian, 2004; Ferson et al., 2006)	1996-2015	US, EU	SRI, Environment	negative	small	countries: value (EU), n.s.s. (US)	countries: positive (EU), n.s.s. (US)
Soler-Domínguez et al. (2015)	Carhart (1997)	-	2002-2013	US	SRI	subperiods: positive (crisis), negative (non-crisis)	-	-	-

Table 16 - Individual findings of 16 papers at the stock level

Source: own elaboration. Methodology shows the main models, regressions, performance measures and other approaches used by the authors. Modifications/Variables show, respectively, modifications made to the base models and the explanatory variables used in the return regressions. Period refers to the to the longest time period covered by a study. Countries mention the individual countries which were analysed, except for studies that cover many different ones, in which case we refer to regions instead (for e.g. Europe, North America, etc.). We also refer if studies analysed aggregate (SRI) or individual dimensions. Performance difference typically refers to the difference in risk-adjusted performance between SRI Vehicles and Conventional Investments. Size, Value and Momentum refer to biases detected using Fama-French based models (Fama-French, 1993 and Carhart, 1997) or other approaches. Despite using Fama French based models, some authors don't report risk factor loadings. For sake of simplicity we only considerer three categories of results: positive and statistically significant; not-statistically significant; negative and statistically significant. Typically, significance levels range between 1% to 10%. Other abbreviations: F-F (Fama-French); RAP (Risk-adjusted performance); ER (Excess return); AR (Absolute Return); MRI (Morally Responsible Investments); SRI (traditional SRI Investments); misc. (miscellaneous); n.s.s. (non-statistically significant).

Author	**Methodology**	**Modifications/ Variables**	**Period**	**Regions**	**Rating Agency**	**Dimensions**	**Performance difference**	**Size**	**Value**	**Momentum**
Auer et al. (2016)	Carhart (1997), Sharpe Ratio	-	2004-2012	US, EU, AP	Sustainalytics	Environment, Social, Governance, Company	countries: n.s.s. (AP, US), negative (EU and certain industries)	-	-	-
Brammer et al. (2006)	Carhart (1997),	-	2002-2005	UK	EIRIS	Environment, Employee, Community	dimensions: n.s.s. (Environment, Employee, Community); Aggregate: n.s.s.	n.s.s.	n.s.s.	n.s.s.
Brammer et al. (2009)	Carhart (1997), AR Regression (1)	(1) SRI scores, Fundamentals	2000-2004	US	-	SRI	negative	large	growth	-

Author	Methodology	Modifications/ Variables	Period	Regions	Rating Agency	Dimensions	Performance difference	Size	Value	Momentum
Derwall et al. (2005)	CAPM (1), Carhart (1997) (1)	(1) Industry Adjusted: Principal Components Analysis	1997-2003	US	Innovest Strategic Value Advisors	Environment	positive	industry-adjusted: ns (not i.a., large (i.a.)	growth (2000-2003)	n.s.s.
Diltz (1995)	CAPM, statistical tests	-	1989-1991	US	CEP	Environment, Nuclear, Military, Charitable, Animals, Women, Information, South Africa	dimensions: positive (Nuclear, Military, Military Envolvement), n.s.s. (most)	-	-	-
Dravenstott and Chieffe (2011)	AR Regression (1)	(1) Sectors, other factors, social screens	2001-2005	US	KLD	Community, Governance, Diversity, Employment, Environment, Product, Exclusion	negative	-	-	-
Edmans (2011)	Carhart (1997) (1,2)	(1) Industry Adjusted: excess return (2) Characteristics adjusted: excess return	1984-2009	US	-	SRI	positive	weights: small (equal weighted), large (value weighted)	weights: n.s.s. (equal weighted), growth (value weighted)	negative

Author	Methodology	Modifications/ Variables	Period	Regions	Rating Agency	Dimensions	Performance difference	Size	Value	Momentum
Erragragui and Revelli (2016)	Carhart (1997) (1), Smith, Tito, Sharpe, eSDAR	(1) Industry Adjusted: Principal Components Analysis	2007-2011	US	KLD	Governance, Environment, Employee, product, Diversity, Community, Human Rights, Family	dimensions/strategy: n.s.s. (most) positive (governance and engagement) negative (community, human rights and disengagement)	dimensions: small (community, human rights, disengament), large (governance, engagement)	dimensions: no pattern	n.s.s.
Filbeck et al. (2009)	CAPM, Carhart (1997), Sharpe, Treynor, BHAR, event study	-	2000-2007	US	-	SRI	misc.: positive (entry into the list, vs S&P), n.s.s. (long run, vs matched companies)	-	-	-
Galema et al. (2008)	Carhart (1997), Book-to-Market Regressions; ER Regressions (1)	(1) Beta, SRI Score, Fundamentals (Fama and Macbeth, 1973).	1992-1996	US	KLD	Governance, Environment, Product, Diversity, Community, Employee relations	dimensions: n.s.s. (most) positive (community)	dimensions: small (product, governance), n.s.s. (environment, community), large (diversity, employee)	growth	dimensions: positive (employee), n.s.s.
Guerard (1997)	AR Regressions (1)	(1) value, growth (IBES)	1987-1994	US	-	SRI	n.s.s.	-	-	-

Author	Methodology	Modifications/ Variables	Period	Regions	Rating Agency	Dimensions	Performance difference	Size	Value	Momentum
Hill et al. (2007)	CAPM	-	1995-2005	US, EU, Asia	-	SRI	countries: positive (EU) n.s.s. (US, Asia)	-	-	-
Humphrey (2012)	Carhart (1997) (1,2)	(1) Industry Adjusted: Principal Components Analysis (2) idiosyncratic risk-mimicking portfolio	2002-2010	World	SAM	SRI	n.s.s.	models: large, none (industry and idyosincratic risk adjusted regressions)	none	n.s.s.
Kempf and Osthoff (2007)	Carhart (1997)	-	1991-2003	US	KLD	Environment, Community, Diversity, Employee, Product, Human Rights	dimensions: positive (employee, community), n.s.s. (diversity, human rights, product, environment) aggregate: positive	dimensions: small (aggregate, environment, product) n.s.s. (employee, human rights) large (negative, community, diversity)	dimensions: growth (aggregate, most criteria) n.s.s. (human rights, negative)	dimensions: positive (human, negative) n.s.s. (diversity, employee) negative (aggregate, community, environment, product)
Lins et al. (2016)	Carhart (1997) RAP Return Regressions (1)	(1) Fundamentals, Industry adjusted: dummies, Idiosyncratic risk, Carhart (1997) risk loadings	2008-2009	US	KLD/MSCI ESG	SRI	positive	-	-	-

Author	Methodology	Modifications/ Variables	Period	Regions	Rating Agency	Dimensions	Performance difference	Size	Value	Momentum
Mollet and Ziegler (2014)	Carhart (1997)	-	1998-2009	US, EU	ZBK	SRI	countries/subperiods: EU: n.s.s. (1998-2003), negative (2003-2009); US: n.s.s.	countries: large (EU), n.s.s. (US)	subperiod: growth (1998-2003), n.s.s.	misc.: negative (US, 2003-2009)
Trinks et al. (2015)	CAPM, Carhart (1997)	-	1991-2012	US	-	SRI	negative	-	-	-

Table 17 - Individual findings of 16 papers at the index level

Source: own elaboration. Methodology shows the main models, regressions, performance measures and other approaches used by the authors. Modifications/Variables show, respectively, modifications made to the base models and the explanatory variables used in the return regressions. Period refers to the to the longest time period covered by a study. Countries mention the individual countries which were analysed, except for studies that cover many different ones, in which case we refer to regions instead (for e.g. Europe, North America, etc.). We also refer if studies analysed aggregate (SRI) or individual dimensions. Performance difference typically refers to the difference in risk-adjusted performance between SRI Vehicles and Conventional Investments. Size, Value and Momentum refer to biases detected using Fama-French based models (Fama-French, 1993 and Carhart, 1997) or other approaches. Despite using Fama French based models, some authors don't report risk factor loadings. For sake of simplicity we only considerer three categories of results: positive and statistically significant; not-statistically significant; negative and statistically significant. Typically, significance levels range between 1% to 10%. Other abbreviations: F-F (Fama-French); RAP (Risk-adjusted performance); ER (Excess return); AR (Absolute Return); MRI (Morally Responsible Investments); SRI (traditional SRI Investments); misc. (miscellaneous); n.s.s. (non-statistically significant).

Author	**Methodology**	**Modifications/ Variables**	**Period**	**Indices**	**Regions**	**Dimensions**	**Performance difference**	**Size**	**Value**	**Momentum**
Belghitar et al. (2014)	CAPM, Carhart (1997), MSCD, Sharpe, Treynor	-	2001-2010	FTSE4Good	World, US, UK, EU	SRI	negative	countries: small (World, US) large (UK, EU)	-	-
Charfeddine (2016)	CAPM, Sharpe	-	2004-2011	FTSE4Good, DJSI, DSI 400, DJ Islamic Market, FTSE Shariah	World, US, UK	SRI, Islam	negative	-	-	-
Collison et al. (2008)	CAPM	-	1996-2005	FTSE4Good	World, US, UK, EU	SRI	positive	-	-	-
Consolandi et al. (2009)	CAPM, event study	-	2001-2006	DJSSI	World	SRI	n.s.s.	large	-	-

Author	Methodology	Modifications/ Variables	Period	Indices	Regions	Dimensions	Performance difference	Size	Value	Momentum
Kurtz and DiBartolomeo (1999)	CAPM, Fundamental Model, APT (1)	(1) Macroeconomic exposures	1990-1999	KLD400	US	SRI	subperiods: positive (1990-1999); n.s.s. (1991-1995, 1995-1999)	-	growth	-
Kurtz and DiBartolomeo (2011)	CAPM, Fundamental Model	-	1992-2010	KLD400	US	SRI	subperiods: positive (1992-1999); negative (1999-2010)	-	growth	-
Lee and Faff (2009)	F-F (1993) (1)	(1) three momentum factors (Scowcroft and Sefton, 2005)	1998-2002	DJSI	World	SRI	-	large	value	misc.: no pattern (country & industry & stock)

Author	Methodology	Modifications/ Variables	Period	Indices	Regions	Dimensions	Performance difference	Size	Value	Momentum
Lesser et al. (2014)	F-F (1993) (1,2,3)	(1) time-varying beta: conditional approach (Ferson and Schadt, 1996) (2) q-theory (Hou et al., 2015) (3) quality (Asness et al., 2013)	2003-2012	Green indices, DJSI, E. Capital Ethical Global, ESPI, ESI, Global Challenges, HSBC, MSCI,	World	SRI, Green	misc.: positive (green & 2003-2007; negative (SRI & 2008-2012)	dimensions: small (green) n.s.s. (sri)	n.s.s.	n.s.s.
Managi (2012)	Market Switching Model (Hamilton, 1989)	-	2001-2008	DJSI, FTSE4Good, MS-SRI	US, UK, Jap	SRI	n.s.s.	-	-	-
Ortas et al. (2012)	CAPM (1)	(1) time-varying alpha and beta: State space market model (Brooks et al., 1998; Faff et al., 2000; Holmes and Faff, 2004)	2005-2010	BCSI	Brasil	SRI	positive	-	-	-
Ortas et al. (2013)	CAPM (1)	(1) time-varying alpha and beta: State space market model (Brooks et al., 1998; Faff et al., 2000; Holmes and Faff, 2004)	2003-2011	DJSI-AP	Asia Pacific	Environment	n.s.s.	-	-	-
Rehman et al. (2016)	CAPM, CAPM (1), Sharpe Ratio	(1) macroeconomic factors	2002-2014	DJSI, MSCI	Asia	SRI	n.s.s.	-	-	-

Author	Methodology	Modifications/ Variables	Period	Indices	Regions	Dimensions	Performance difference	Size	Value	Momentum
Sauer (1997)	CAPM	-	1986-1994	DSI400	US	SRI	n.s.s.	-	-	-
Schröder (2007)	CAPM, F-F (1993), Sharpe Ratio	-	1992-2003	FTSE4Good, DJSI, Humanix, KLD, NAI, West, Jantzi, Kempen, Ethical, Aspi, Calvert	World, US, Canada, EU, Sweden, Australia	SRI	n.s.s.	small (NAI)	indices: growth (DSI400); value (NAI)	-
Statman (2006)	F-F (1993) (1), Sharpe, Alpha-s (Statman, 1986)	(1) Industry adjusted: industry weighted indices	1990-2004	DSI 400, Calvert, Citizens, DJSI	US	SRI	n.s.s.	large (DSI400)	growth (DSI400)	-
Wu et al. (2017)	ANOVA, CAPM, Sharpe	-	2004-2011	FTSE4GOOD	UK	SRI	positive	-	-	-

Table 18 - Individual findings of 7 papers at more than one level

Source: own elaboration. Methodology shows the main models, regressions, performance measures and other approaches used by the authors. Modifications/Variables show, respectively, modifications made to the base models and the explanatory variables used in the return regressions. Period refers to the to the longest time period covered by a study. Countries mention the individual countries which were analysed, except for studies that cover many different ones, in which case we refer to regions instead (for e.g. Europe, North America, etc.). We also refer if studies analysed aggregate (SRI) or individual dimensions. Performance difference typically refers to the difference in risk-adjusted performance between SRI Vehicles and Conventional Investments. Size, Value and Momentum refer to biases detected using Fama-French based models (Fama-French, 1993 and Carhart, 1997) or other approaches. Despite using Fama French based models, some authors don't report risk factor loadings. For sake of simplicity we only considerer three categories of results: positive and statistically significant; not-statistically significant; negative and statistically significant. Typically, significance levels range between 1% to 10%. Other abbreviations: F-F (Fama-French); RAP (Risk-adjusted performance); ER (Excess return); AR (Absolute Return); MRI (Morally Responsible Investments); SRI (traditional SRI Investments); misc. (miscellaneous); n.s.s. (non-statistically significant).

Author	Methodology	Modifications/ Variables	Period	Regions	Level of analysis	Dimensions	Performance difference	Size	Value	Momentum
Blanchett (2010)	Carhart (1997), statistical tests	-	1990-2008	World	funds, indices	SRI	n.s.s.	indices: n.s.s. (Calvert, FTSE4GOOD US), large (DSI400); funds: n.s.s. (average coef.) large (weighted coef.)	indices: n.s.s. funds: n.s.s. (average coef) value (weighted coef.)	indices: n.s.s. funds: n.s.s. (average coef) value (weighted coef.)
Cortez et al. (2009)	CAPM (1)	(1) time-varying alpha and beta: conditional approach (Christopherson et al., 1998)	1996-2007	UK, EU	funds, indices	SRI	n.s.s.	-	-	-
Scholtens (2007)	F-F (1993)	-	2001-2005	Netherlands	funds, indices	SRI	n.s.s.	indices: large; funds: no pattern	indices: growth; funds: n.s.s.	-
Schröder (2004)	CAPM (1,2,3), Sharpe Ratio	(1) blue-chip and small stock benchmarks (2) Timing (Treynor and Mazuy, 1966)	1990-2002	US, Germany, Sweden	funds, indices	SRI	n.s.s.	funds: small (Germany, Sweden), large (US)	-	-

Author	Methodology	Modifications/ Variables	Period	Regions	Level of analysis	Dimensions	Performance difference	Size	Value	Momentum
		(3) time-varying beta: conditional approach (Ferson and Schadt, 1996)								
Statman (2000)	CAPM, eSDAR	-	1990-1998	US	funds, indices	SRI	n.s.s.	-	-	-
Vermeir (2005)	CAPM, F-F (1993), Information ratio	-	1998-2004	World, US, EU	stocks, indices	Human Resources, Corporate Governance, Society, Environment, Clients and Suppliers,	n.s.s.	dimensions: large (clients and supplier) n.s.s. (remaining) indices: large (most) n.s.s. (DSI)	dimensions: growth (clients and suppliers, governance), value (society), n.s.s. (environment, human resources) indices: n.s.s. (most) growth (DSI, Ethibel Global)	-

Author	Methodology	Modifications/ Variables	Period	Regions	Level of analysis	Dimensions	Performance difference	Size	Value	Momentum
Xiao et al. (2015)	ICAPM (Merton, 1973), CAPM, F-F (1993), Carhart (1997)	-	1990-2013	US	funds, indices	SRI	models: n.s.s. (ICAPM), negative (traditional), positive (DSI400)	-	-	-

Table 19 - Identification of Crisis Periods.
Source: Own elaboration. We summarise the various approaches chosen by authors in literature to identify the three major crisis periods that occured since the early 2000's. For sake of simplicity, in this table we only refer to authors that have explicitly identified and analysed a whole crisis period. Sources include the Economic Data Base of the Federal Reserve of St. Louis (FRED St. Louis), European Central Bank (ECB) and specific methods devised by the authors themselves or by other researchers.

Author (Year)	**Dot Com Crisis**	**Global Financial Crisis**	**Sovereign Debt Crisis**	**Method/Source**
Amenc and Le Sourd (2010)		Jan. 2007 - Dec. 2009		none
Becchetti et al. (2015)	Mar. 2001 - Nov. 2001	Dec. 2007 - Jun. 2009		FRED (St. Louis)
Derwall et al. (2005)	Mar. 2000 - Dec. 2003			none
Hooi et al. (2015)		Sep. 2008 - May 2009		Ang (2015); Lean and Nguyen (2014)
Leite et al. (2015)	Jan. 2001 - Mar. 2003	Jun. 2007 - Feb. 2009	May 2011 - May 2012	Pagan and Sossounov (2003)
Lesser et al. (2016)	Mar. 2000 - Oct. 2002	Oct. 2007 - Mar. 2009		Peaks and troughs MSCI AC World Index
Lins et al. (2016)		Aus. 2008 - Mar. 2009		Lins, Volpin and Wagner (2013)
Muñoz et al. (2014)	Mar. 2000 - Oct. 2002	Oct. 2007 - Mar. 2009	Oct 2009 - Jan. 2013	ECB
Nofsinger and Varma (2014)	Mar. 2001 - Nov. 2001	Dec. 2007 - Jun. 2009		NBER
Ortas et al. (2013)		Jan. 2008 - Jan. 2009		none
Soler-Domínguez and Matallín-Sáez (2015)		Jan. 2008 - Jun. 2009		NBER
Varma and Nofsinger (2012)	Mar. 2000 - Oct. 2002	Oct. 2007 - Mar. 2009		Peaks and troughs S&P

Table 20 - Dimensions definitions

Source: own elaboration. We present the definition of dimensions according to some of the most used rating agencies we found in the literature. They are based on summaries presented by some papers that analysed them, namely: KLD (Galema et al., 2008); Vigeo (Vermeir, 2005); Sustainalytics (Auer et al., 2016); EIRIS (Brammer et al., 2006).

Dimensions	Agency/Source	Description
Client and Supplier	Vigeo	Considers the quality of management of client and supplier relationships.
Community	EIRIS	Indicator based on one measure: community responsiveness.
Diversity	KLD	Deals with the composition of the workforce, especially senior management and the board.
	Sustainalytics	Comprise issues such as labour relations, business practises, employment diversity and community involvement.
Eco-eficiency	Innovest	Value that a company generates relative to the waste it requires to do so.
Employee Relationships	KLD	Deals with the relationship between employees and the company and with especially concern issues related to employee compensations.
Employee Responsability	EIRIS	Indicator based on six measures: health and safety systems, systems for employee training and development, equal opportunities policies and systems, systems for good employee relationships, systems for job creation and security.
Environment	EIRIS	Indicator based on three measures: quality of environmental policies, management systems and reporting.
	KLD	Environment management and policies.
	Sustainalytics	Measures Environment involvement though proactive initiatives such as recycling, waste reduction, biotechnology, environment cleanup and renewable energies.
	Vigeo	Considers the way the company affects the environment through its activities.
Governance	KLD	Deals with how the firm is governed and directed but resulted from the renaming of the category "other" in 2002.
	Sustainalytics	Deals with outstanding best practises related to Board independence and elections, Auditor independence, executives' compensations, voting and shareholder voting rights.
	Vigeo	Rates the transparency and efficiency of governance in relation to shareholders.
Human Resources	Vigeo	Continuous improvement in employment conditions, evaluation of employee skills and employability.
Product	KLD	Strengths and Weaknesses about the product quality and production processes.
Society	Vigeo	Contributions of the company to the community such as public causes and employee training.

Appendix 2 – Selection Criteria of SRI Indices

In the selection process of SRI Indices, we accounted for certain criteria we deemed relevant based on our findings from the literature. The main criteria and the respective justification are represented in the following table:

Table 21 - List of Criteria and respective justification used in the Selection Process of SRI Indices
Source: Own elaboration

<table>
<tr><th>Criteria</th><th>Justification</th></tr>
<tr><td>ESG Indices
We only selected Indices that used environmental, social and governance (ESG) criteria. We, therefore, excluded indices that focus on just one dimension (for e.g. environment Indices).</td><td>ESG Dimensions represent the most common concerns of SRI. Based on our findings from the literature, we concluded that the analysis of specific dimensions has not proved to be relevant given the disparity of findings in the literature. Moreover, we showed there’s a lack of consensus on the definition of such individual dimensions.</td></tr>
<tr><td>Total Return Indices
We only selected SRI Indices from which it was possible to collect total returns.</td><td>This option is often used in the Literature by authors such as Hill et al. (2007), Kreander et al. (2005), Ortas et al. (2012); Ortas et al. (2013), Schröder (2007), among others.</td></tr>
<tr><td>Official Benchmark
We selected SRI Indices for which data was available for their official benchmark.</td><td>As we have shown in the Review of the Literature, the choice of benchmark can have a tremendous effect on findings. We argue that official benchmarks are the most appropriate option since the only difference between a SRI Index and its official benchmark is the application of ethical filters. Moreover, this option is also often used in the literature (see for e.g. Schröder, 2007; Vermeir et al., 2005).</td></tr>
<tr><td>Series
We selected indices belonging to a Series instead of isolated indices.</td><td>Each Series has several indices with similar time spans that cover different geographical areas which allows a more consistent analysis.</td></tr>
<tr><td>Different Series and Regions
We selected different Series instead of choosing just one.</td><td rowspan="2">Our Review of the Literature showed very heterogenous findings in SRI literature, due to multiple factors. Given the different conceptions of SRI between ratings agencies and the different findings in terms of countries, we avoid sample specific results by choosing more than one source of SRI, not analysing just one region and by the possibility of comparing indices of each region across the three Series.</td></tr>
<tr><td>Similar Regions across Series
We procured to select Series with indices with similar geographic scope</td></tr>
<tr><td>Fama-French Factors Scope
We only selected Indices with a Geographical scope consistent with the regions covered by the factors available at Kenneth French Library. This scope is: Global, Global ex-US, Europe, North America, Japan, Asian-Pacific ex Japan</td><td>We use Fama-French models to analyse SRI Indices and the source we use to retrieve data on each one of the Fama-French factors is Kenneth French Library.</td></tr>
</table>

The application of such criteria resulted in the selection of Indices from the FTSE4GOOD, MSCI ESG and STOXX ESG Series. Unfortunately, we were not able to select indices from other SRI Indices often studied in literature for some reasons related to the application of such criteria. One example are Indices from the Dow Jones Sustainability Series, which we did not include in our analysis due to the fact that we were not able data on Datastream related to Total Returns. Other indices such as Calvert were not included because they were isolated indices and we were not able to obtain data for its official benchmark.

Appendix 3 – Description of Fama-French 5 factors

Table 22 - Description of the geographic scope of Fama-French 5 factors for Developed Countries
Source: Kenneth French Library. Available at:
http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_5developed.html

Country	Global	Europe	Japan	Asia Pacific ex Japan	North America
Australia	✔			✔	
Austria	✔	✔			
Belgium	✔	✔			
Canada	✔				✔
Switzerland	✔	✔			
Germany	✔	✔			
Denmark	✔	✔			
Spain	✔	✔			
Finland	✔	✔			
France	✔	✔			
Great Britain	✔	✔			
Greece	✔	✔			
Hong Kong	✔			✔	
Ireland	✔	✔			
Italy	✔	✔			
Japan	✔		✔		
Netherlands	✔	✔			
Norway	✔	✔			
New Zealand	✔			✔	
Portugal	✔	✔			
Sweden	✔	✔			
Singapore	✔			✔	
United States	✔				✔

Appendix 4 – GMM Estimation Procedure

The Generalized Method of Moments (GMM) estimation procedure starts by assuming the following moment conditions[74] for each equation[75]:

(A.1) $$E(g_i)=E(x_i \cdot \varepsilon_i)=E[x_i \cdot (y_i - z_i' \delta)]=0, \text{ (for all i=1, 2, … , n)},$$

Where n represents that number of observations, x_i is K-dimensional vector of predetermined regressors[76], z_i is a L-dimensional vector of regressors of the original equation and δ is a L-dimensional vector of parameters. The procedure implies substituting the moment conditions by their sample analogue:

(A.2) $$g_n=\frac{1}{n}\sum_{i=1}^{n} g_i = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot (y_i - z_i' \hat{\delta}) = 0, \text{ (for all i=1, 2, … , n)},$$

Where $\hat{\delta}$ is a L-dimensional vector of estimates of δ. Since there are K predetermined variables, there are K moment conditions. The moment conditions of a GMM system are just a collection of the individual moment conditions of each equation. Each one represents a restriction that must be satisfied by the estimates, which implies that the system of equations is fully identified when the number of restrictions is equal to the number of parameters to be estimated, i.e., when K=L. This implies that is possible to find a vector of estimates $\hat{\delta}$ that satisfied each moment condition. However, when K>L, the system is said to overidentified and no exact solution can be found. The GMM estimator is the vector $\hat{\delta}$ that can make the sample moments as close to zero as possible and results of minimizing the following objective function:

(A.3) $$J(\hat{\delta}, \widehat{W}) = n \cdot g_n'(\hat{\delta}) \cdot \widehat{W} \cdot g_n(\hat{\delta})$$

---

[74] These conditions are also called orthogonality conditions since they imply that expected product of two variables is zero.

[75] The following description of the GMM estimation procedure is adapted from Hayashi (2000).

[76] We use the definition of Hayashi (2000) and define predetermined regressors as regressors that are orthogonal to the error term ε_i.

Where $\widehat{W}$ is a positive semi-definite random Weighting Matrix which serves to weight the different moment conditions when constructing the distance measure. $J(\hat{\delta}, \widehat{W})$ is called the J-statistic. The value of this function is only zero if no instrumental variables are used. However, imposing any additional constraint causes the system to be overidentified and the J-statistic to have the following distribution:

$$J(\hat{\delta}, \widehat{W}) \sim x^2(K - L)$$

Therefore, the J-statistic can be used to perform a test of overidentifying restrictions (Hansen, 1982). This test analyses the null hypothesis that a parameter is zero throughout the system:

$$\text{H0: } \theta_{\mathrm{i,j}} = 0$$

$$\text{H1: } \exists\, i, j : \theta_{i,j} \neq 0$$

Where $\theta_{\mathrm{i,j}}$ represents some coefficient j from the equation of Index i. For example, an estimation of the Fama-French (1993) model for the FTSE4GOOD Series implies requires 28 orthogonality conditions since there are 4 regressors per equation and 7 equations. In the absence of any restraints, the system is fully identified and the J-statistic is zero since there are 28 parameters to be estimated. However, restricting alpha to be zero throughout the System implies reduces parameters to be estimated to 21. Since the system is overidentified, the J-statistic is no longer zero and follows a chi-square distribution with 7 degrees of freedom.

Appendix 5 – Test of overidentifying Restrictions in Eviews 9

We used the econometric software Eviews 9 to estimate the various models we used in a GMM-System framework. This software can produce the test of overidentifying restrictions for individual equations but unfortunately not for a System of equations. However, it is possible to manually perform it.

As an illustrative example, we show an example of how to obtain the J-statistic associated with SMB for the case of the Fama-French (1993) model applied in a GMM System to the FTSE4GOOD Indices. As we have previously explained, this test statistic serves to test the null hypothesis that the estimates of a regressor are null throughout a system of equations. In this example, this test-statistic will test the null hypothesis that estimates of SMB Factor are null throughout the system of Fama-French (1993) equations applied to FTSE4GOOD Indices.

To obtain this statistic, it's necessary to construct another System, which is specified in the table below. In the original system, since there are 7 regressions with 4 regressors each, there are 28 restrictions and parameters to be estimated, and the system is exactly identified. In this new system, the SMB regressor is excluded from each equation but is identified as an instrument by the "@" command, therefore restricting it to be zero in each of the 7 regressions, decreasing the number of parameters to be estimated to 21. Thus, this procedure creates conditions to perform a test of overidentifying restrictions, since the resulting system is overidentified and the resulting J-statistic follows a chi-square distribution with 7 degrees of freedom. The procedure is very similar when applied to other regressors and models.

Table 23 - Example of a specification of a GMM System of the Fama-French (1993) model applied to FTSE4GOOD indices used to apply the Sargan (1984) test to the SMB factor

Source: Adapted from econometric software Eviews 9. We highlight in bold the fact that the SMB regressors are identified as instruments in each equation.

FTSE4GOOD_EUR_50_RF=C(1)+C(2)*ftse_dev_eur_rf+C(4)*eur_hml @ c ftse_dev_eur_rf **eur_smb** eur_hml
FTSE4GOOD_EUR_RF=C(5)+C(6)*ftse_dev_eur_rf+C(8)*eur_hml @ c ftse_dev_eur_rf **eur_smb** eur_hml
FTSE4GOOD_GLB_100_RF=C(9)+C(10)*ftse_dev_rf+C(12)*glb_hml @ c ftse_dev_rf **glb_smb** glb_hml
FTSE4GOOD_GLB_RF=C(13)+C(14)*ftse_dev_rf+C(16)*glb_hml @ c ftse_dev_rf **glb_smb** glb_hml
FTSE4GOOD_JAP_RF=C(17)+C(18)*ftse_jap_rf+C(20)*jap_hml @ c ftse_jap_rf **jap_smb** jap_hml
FTSE4GOOD_US_100_RF=C(21)+C(22)*ftse_us_rf+C(24)*na_hml @ c ftse_us_rf **na_smb** na_hml
FTSE4GOOD_US_RF=C(25)+C(26)*ftse_us_rf+C(28)*na_hml @ c ftse_us_rf **na_smb** na_hml

Appendix 6 – List of industries included in the PCA

Table 24 - List of Industries included in the Principal Components Analysis
Source: Adapted from Kenneth French Library - Available at:
http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_49_ind_port.html

Agriculture	Rubber and Plastic Products	Coal	Retail
Food Products	Textiles	Petroleum and Natural Gas	Restaurants, Hotels, Motels
Candy & Soda	Construction Materials	Utilities	Banking
Beet & Liquor	Construction	Communication	Insurance
Tobacco Products	Steel Works	Personal Services	Real Estate
Recreation	Fabricated Products	Business Services	Trading
Entertainment	Machinery	Computers	Other
Printing and Publishing	Electrical Equipment	Computer Software	
Consumer Goods	Automobiles and Trucks	Electronic Equipment	
Apparel	Aircraft	Measuring and Control Equipment	
Healthcare	Shipbuilding, Railroad Equipment	Business Supplies	
Medical Equipment	Defense	Shipping Containers	
Pharmaceutical Products	Precious Metals	Transportation	
Chemicals	Non-Metallic and Industrial Metal Mining	Wholesale	

Appendix 7 – Annualization of Alpha

Many authors in the SRI literature that have used single and multi-factor models annualized the alpha's obtained from the respective models[77]. Thus, in coherence with extant literature we follow the same procedure. We annualize using each alpha using a standard formula:

$$\alpha^{annual} = 100 * [(1 + \frac{\alpha^{daily}}{100})^{252} - 1] \tag{A.4}$$

Where α^{daily} represents the original daily alpha, α^{annual} the annualized alpha we present in the results and 252 the number of trading days for which SRI Indices have observations within a year.

[77] See for example Derwall et al. (2005), Erragragui and Revelli (2016), Galema et al. (2008), among others.